Expert Finding in Disparate Environments

Raymond J. D'Amore

March 2008

Submitted in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy
in the Department of Information Studies
at the University of Sheffield, Sheffield, United Kingdom

Abstract

Providing knowledge workers with access to experts and communities-of-practice is central to expertise sharing, and crucial to effective organizational performance, adaptation, and even survival. However, in complex work environments, it is difficult to know *who knows what* across heterogeneous groups, disparate locations, and asynchronous work. As such, where expert finding has traditionally been a manual operation there is increasing interest in policy and technical infrastructure that makes work *visible* and supports automated tools for locating expertise.

Expert finding, is a multidisciplinary problem that cross-cuts knowledge management, organizational analysis, and information retrieval. Recently, a number of expert finders have emerged; however, many tools are limited in that they are extensions of traditional information retrieval systems and exploit artifact information primarily. This thesis explores a new class of expert finders that use organizational context as a basis for assessing expertise and for conferring trust in the system. The hypothesis here is that expertise can be inferred through assessments of work behavior and work derivatives (e.g., artifacts).

The *Expert Locator*, developed within a live organizational environment, is a model-based prototype that exploits organizational work context. The system associates expertise ratings with expert's signaling behavior and is extensible so that signaling behavior from multiple activity space contexts can be fused into aggregate retrieval scores. Post-retrieval analysis supports evidence review and personal network browsing, aiding users in both *detection* and *selection*. During operational evaluation, the prototype generated high-precision searches across a range of topics, and was sensitive to organizational role; ranking *true* experts (i.e., *authorities*) higher than *brokers* providing referrals. Precision increased with the number of activity spaces used in the model, but varied across queries. The highest performing queries are characterized by high specificity terms, and low organizational diffusion amongst retrieved experts; essentially, the highest rated experts are situated within organizational niches.

2

Acknowledgements

If this thesis is a journey's end then I can only start by thanking those who made the first steps possible. Thanks to Professor Peter Willet who provided the initial opportunity to study at Sheffield. Of course, the actual work would not have been possible without the continual support of my advisors: Professor Micheline Beaulieu and Dr. Mark Sanderson. They were both instrumental in guiding me from near and afar and most importantly provided the right measure of oversight and critical review. As my primary supervisor, Professor Beaulieu facilitated initial entry into the program in time to meet a rapidly shrinking admissions window, and provided guidance through various program milestones. Micheline and Mark thank you both for all your support; especially during times when certain events would have otherwise made the continuation of this work less likely. In addition, thanks go to my two examiners Professor Steve Whittaker and Dr. Arjen de Vries who provided useful feedback and an interesting Viva.

This work was carried out at the MITRE Corporation with partial support from internal research funding. However, much was done outside the context of an actual project and so I must thank my supervisors Ms. Julie Gravallese and Dr. Mark Maybury for their internal support for this research to include work accommodations that were essential to the completion of this research. I was indeed fortunate to have had a number of colleagues that shared interest in this work or supported it in many ways to include Manu Konchady, a close colleague with whom a number of ideas on many topics were discussed; Paul Lehner for always asking me "why are you doing this?" which I think was a type of encouragement—it was, wasn't it Paul? Sue Sanders who was there to help with documents, plan travel to Sheffield, and simply be a friend. However, the biggest thanks goes to Dan Helm and Gary Valley, close colleagues and closer friends, who were always there to listen to a new idea, talk about the next model, crack a joke, but who also provided programming support for the operational prototype.

With this thesis a pledge is fulfilled; a promise delivered. For me it serves as a thank you and expression of love for my parents Stephen and Sandina D'Amore who did nothing if not love their family. Thanks to my three sons Christopher, Ryan, and Kyle who were most supportive and encouraging and to our new daughters-in-law, Conny and Victoria. Special thoughts are for Mason who has given our family new joy, my Mother-in-law Dawn for her continual interest and support, and to my sister Marie and her family for the closeness we have.

But the one who most made this possible is my wife Ellen. Ellen you stood by me, gave me support, and kept me going even when it would have been easy to put this aside. Your love never diminished throughout; without you this would not have been possible. To you, this thesis is dedicated!

Table of Contents

List of Figures

List of Tables

1 Introduction

Experts are critical to organizational success; collectively they serve as cross-organizational linchpins tying together otherwise narrowly channeled groups. Experts serve as consultants, mentor staff, and embody elements of corporate memory through work artifacts and storytelling. Within technical organizations, such as The MITRE[1] Corporation, experts take on central roles in defining research directions, assessing research proposals, and monitoring work. For example, MITRE's Technology Area Teams (TATs) are cross-organizational groups consisting of expert technologists with proven track records in research, applications, and program development. TATs play a key role in developing research roadmaps and in assessing both internal and external research relevant to MITRE's business areas. More generally, experts are situated within a particular work setting taking on formal and informal roles that are shaped by work domain and culture. Outside the traditional enterprise, experts take on long-standing roles such as consulting to news agencies, testifying in legal proceedings, advising on environmental issues, and providing help within virtual communities. The need to find experts is not bound to a particular setting.

Yiman-Seid and Kobsa (2003) identified a number of reasons for locating experts to include problem definition, assessment and analysis, information filtering, and project tasking. However, in large heterogeneous environments, expertise location is problematic. Experts are often difficult to find due to widely varying work contexts, disparate locations, and asynchronous work. The problem is exacerbated by work compartmentalization where tasking is shielded to comply with privacy or need-to-know restrictions. In complex work environments, it is difficult to know *who knows what*. As such, where expert finding has traditionally been largely a manual operation there is increasing interest in policy and technical infrastructure that makes work *visible* and supports automated tools for locating expertise.

---

[1] www.mitre.org Accessed on July 24, 2007

Expert finding, is a multidisciplinary problem that cross-cuts knowledge management, organizational analysis, and information retrieval. More recently, a number of tools to support expert finding have emerged, for example, Yimam, (1999), and TREC Enterprise Track (TRECENT), Craswell et al (2005). For the most part, these tools are limited in that they are simple extensions of information retrieval and knowledge management systems and typically exploit a single source of information, for example, e-mail. The premise is that expertise can be inferred simply by counting up relevant documents (e.g., e-mail posts or publications). As such current systems may not reflect characteristics of real experts or align with organizational structure and work behavior. Just as automated retrieval systems address relevance in the context of a collection, expert finders need to use organizational context to assess expertise.

The goal here is to explore a new class of expert finders that use organizational context as a basis for assessing expertise and for conferring trust in the system. The hypothesis is that expertise can be inferred through assessments of work behavior and work derivatives (e.g., artifacts) and that system trust or reliability can be conferred by embedding experts in their personal networks. Personal networks which subsume work activities, organizational ties, and artifacts provide users with context needed to discern true experts from those that may simply have an interest in a topic. While machine-generated personal networks may, at best, be approximations of actual personal networks maintained by individuals, the notion here is that automatically constructed personal networks will provide organizational context useful in assessing whether the *expert* is really an expert.

Experts are critical to creating organizational value Huber (1999). While there is debate on just what constitutes expertise, there is general agreement that expertise is situated; it depends on work context, organizational culture, and human judgment. As such, the issue of "what is expertise" is best viewed in the context of a target environment. McDonald and Ackerman (1998) defined expertise as "the embodiment of knowledge and skills within individuals." Others have operationalized expertise to fit a particular domain; for example, Maybury, D'Amore, House (2003) describe expertise in the context of the MITRE Corporation as "knowledge of MITRE's mission and sponsor program areas coupled with specific technical, management, and business skills needed to support clients and conduct research." This definition

is *actionable*; it provides a context for thinking about how expertise is exchanged or signaled within the enterprise.

Expertise sharing enables team formation and community emergence. Knowing the skills or experiences of potential team members is important in ensuring effective resource utilization in formal and informal tasking. While to some extent expertise sharing has been subsumed into knowledge management (KM) initiatives, most organizations have focused on building knowledge stores and technical means for accessing artifacts. More recently, however, the focus has shifted towards managing expertise and this has centered on knowing "who knows what" and "who works with whom". For example, at MITRE, "…the goal of KM is not to capture everything that people know, but rather to create an environment that fosters knowledge exchange, capture, reuse, and internalization."[2]

However, organizations often don't know what they know Hinds and Pfeffer, (2003). While it is in part due to the specialized nature of expertise, cognitive and motivational constraints also contribute. For example, in many environments, people compete for particular roles, formal positions, funding, and promotions. Individuals are sometimes rewarded even though the work is supported by a team. Competition may act to curtail cooperation by inhibiting trust formation. The work environment may also make it difficult to share or signal expertise. In organizations with a strong mission/market orientation, work may become "stovepiped" or compartmented inhibiting expertise and knowledge sharing. Staff may work in geographically disparate or transient environments and this may reduce communication with others or may limit their ability to make visible the kinds of work they are doing; this includes staff members who work at remote sites, telecommute, or work in mobile environments. Others may be outwardly focused on external communities, government organizations, industry, or academia and may have little connection with the main work of the enterprise. Interestingly, with visibility comes responsibility. Acknowledged experts may be *asked* to take on roles as mentors, to answer questions, or to provide help without having any formal support for the job. While the expert may be expected to provide a service he/she may not have the resources or formal support needed to take on the role of *advisor*. There may be concern that providing wrong answers may

---

[2] Quote from an interview with then MITRE CIO Al Grasso, 2000.

incur certain career risk and poor performance may surface at annual performance review. As such, the problem of detecting expertise is exacerbated by organizational constraints imposed on how people work, interact with others, and signal their expertise. This motivates research in an operational setting so as to address some of these issues; the setting here is the MITRE Corporation.

1.1   Research Environment: The MITRE Corporation

> *"The MITRE Corporation is a not-for-profit organization chartered to work in the public interest. As a national resource, we apply our expertise in systems modernization to address our sponsors' critical needs. MITRE has 6,500 scientists, engineers and support specialists—65 percent of whom have Masters or Ph.D. degrees. Staff members work on hundreds of different projects across the company, demanding a high level of technical, operational, and domain knowledge." [3]*

MITRE is a knowledge-based organization and, as shown in Figure 1-1, MITRE's knowledge management (KM) initiative culls out processes for capture/reuse of knowledge, cultural influences, and enabling technologies needed to facilitate knowledge sharing and information exchange.   MITRE is engaged in community development; environments that provide a context for creating new work groups, and for distributing expertise across organizational boundaries more efficiently.

Figure 1-1 Knowledge Management: An Enterprise Perspective, Small and Zoracki (2000)

---

[3] http://www.mitre.org/about/ MITRE Mission Statement, Accessed on July 25, 2007

MITRE is a rich environment for conducting expert finder research since many of the knowledge-based services support expertise sharing at some level. Current services evolved from a function-based statement of needs that takes a user view on business and information technology requirements. These needs relate to *expertise management* through a number of perspectives such as *how users will go about doing their job (business process); how users will be enabled to do their job using information technology; how users will work as individuals or members of teams, and how users will interact with sponsors and other external entities.* Most of these needs are based on understanding *who knows what* and *who works with whom* and are therefore at the center of this research.

Expertise sharing is enabled by use of global video teleconferencing, the MITRE Information Infrastructure (MII)[4], as well as public key infrastructure (PKI) enabled extranet services. MITRE has a number of formal expertise management services such as InfoDesk, Technical Area Teams, and Technology Integrators that serve up expertise in established technology or business areas. Staff members can also peruse user Share Folders to find relevant documents made publicly available by authors, and may also use the enterprise search system to locate key artifacts.

MITRE has expertise in a wide range of disciplines and problem domains to include: systems engineering, computer science, natural language processing, air traffic control, biological science, the social sciences, and others. The problem of tracking expertise is especially important given the diverse sponsor base, and mission areas. Employees are dispersed worldwide in line with the national security mission, and geographic disparateness adds to the problem of identifying relevant expertise and supporting effective collaboration. MITRE has introduced a number of services and business practices designed to mitigate the problem of finding expertise. For example, a manually built expertise directory is now used by MITRE's INFODESK to provide users with a points-of-contact directory for locating experts. MITRE's HOTLINE is another way to get support for locating experts although the HOTLINE generally

---

[4] The MII is MITRE's corporate Intranet; it was awarded the CIO Magazine 1999 Enterprise Value Award (EVA). Corporate tools and services discussed here reflect the MII during the time this research was conducted.

provides pointers to specific organizations and it is most useful for tracking down general topics; not highly specialized areas. The MII provides search services for access to published documents that allows users to use authorship and content to get to the needed expertise. Newsgroups provide another way to find expertise based on the newsgroup focus and explicit postings. Users can search across multiple lists, browse posts, or have them emailed to their desktop as part of a current awareness capability. Users may also use the organization chart to find expertise. MITRE centers, divisions, and departments are often bounded by specific skills and sponsor bases so that it is possible, say, to find communication engineers in one or two departments.

The use of skills databases populated manually by knowledge engineers and/or employees is not new and in some organizations are the de facto methods for capturing expertise. At MITRE, a skills database proved difficult to build, and problematic to keep current. Expertise was difficult to capture using manual update mechanisms and users found it difficult to encode both general and specific knowledge; especially in areas where skills were changing rapidly. MITRE's skills database is no longer operational.

While MITRE has built, deployed, and in some cases abandoned a number of methods for expertise location, many rely on their personal network. On aggregate, expert finding services are not well integrated, do not cover many domains or specialties, and require certain "overhead" to use. For example, some services require filling out an online form. To most, canvassing their personal networks is inherently more "user friendly". It is in part a cultural and learned behavior to call those you know in order to find answers to questions, help on a problem, and referrals to experts. As will be discussed later in this thesis, the referral network has limitations related to "anchoring" biases such that local searches of one's personal network may lead to a form of suboptimal convergence and preclude finding experts in disparate parts of the organization. This is especially the case for new employees with limited contacts or those not connected to major work areas. Therefore, while there are a few loosely organized services, telephone, and email support for expert finding, there is no system that scans the corporation on a continuous basis and produces a consolidated view of enterprise expertise.

Expertise is hard to track in dynamic environments. MITRE work is dispersed across a number of work environments; its employees work at fixed locations but may work off-site or are mobile. While the core expertise is centered at the two main campuses (Bedford, MA and Washington, D.C.) specialized knowledge of sponsor environments exists at a number of sites. In addition, a significant percentage of workers telecommute and dial into the corporate network often using low-bandwidth connections. All of this suggests that access to needed expertise is contingent on capturing knowledge from disparate locations and that expertise must be shareable across various communications environments. This includes mobile workers within the *MITRE-footprint* and those that work outside of it. Depending on the characteristics of the mobile device and its location, service level may vary considerably. This is more than a communication problem as it may be difficult to track work crossing organizational boundaries and communication gateways, capture it in some kind of expertise profile, and share it with others. The difficulty of sharing expertise is compounded by the diversity of users, which includes MITRE technical staff, knowledge workers, developers, support staff, legal, human resources, new employees, business partners, and others in the research community. User diversity implies additional constraints imposed regarding information exchange, communications interoperability, and intellectual property or privacy restrictions on sharing.

Expert finding is envisioned as part of an expertise management framework. Expertise detection and sharing is part of everyday work. While initial research may be focused on expert finders as a class of information retrieval system, ultimately expert finding will be embedded in various work contexts such as tools for sponsor and contractor support and access, network appliances, integrated messaging, calendaring and resource scheduling, desktop environments, and workflow management. This does not preclude expert finding as a key task in external, multi-organizational environments as well, Becerra-Fernandez (2000). The mantra here is *access to expertise anywhere by anyone*.

1.2   Problem Focus

This thesis is focused on providing new methods for locating expertise in order to address limitations in current practice. The goal is to develop an expertise locator that can be used with

little if any manual support needed to find individual experts as well as expertise networks, groups of experts with commons skills and work activities. A main objective is to explore the confluence of traditional information retrieval, and social and organizational network analysis. Where information retrieval provides a solid basis for collecting, indexing, and storing artifacts or evidence, social network analysis provides a basis for transforming document lists to organizational networks. In effect, users will assess relevance through a social lens supported by document (artifact) evidence. While the actual prototype must address *detection* it must also incorporate knowledge of the *selection* problem. Selection may require different strategies than detection since choosing which experts to contact or work with may require organizational knowledge or insights from colleagues that may go beyond simply producing a list of candidates. In order to make the research manageable, the core search algorithm and the subsequent operational evaluation focus on detection; however, the prototype has incorporated special features to support selection.

Expert finder evaluation starts with the position that judging document relevance is qualitatively different from assessing expertise. In particular, while document relevance may be a component of an overall assessment, expertise judgments may be formed from other factors related to work context. Assessing whether a person is "relevant" to a topic (i.e., has significant expertise) requires knowledge of a person's activities, interactions with others, and specific roles played within and outside the enterprise. While TREC-like methods that use document pooling strategies provide a guide for establishing relevance sets; they are not easily fitted to operational environments such as encountered at MITRE. Here, it is problematic to a priori specify query-relevant sets (qrels) without significant cost or bias; especially when queries cross-cut multiple disparate domains to include special niche areas not easily assessed by judges with general knowledge. Here, a novel survey-based sampling scheme technique is used to generate a query-specific expertise network which is used as a baseline in which to evaluate *Expert Locator* retrieval results.

1.3   Evaluation Data Archive

The evaluation methodology, Chapters 8 and 9, is written, in part, to promote future investigation into the use of contextualized evidence (i.e., activity spaces) as a basis for identifying expertise. Various evaluation aspects are detailed to include test queries, performance measures and analyses with detailed results down to the individual query and activity space level. Further, special methods are culled as to the underlying formalism, specific instruments used, and experiment protocol followed. For example, an extensive discussion on snowball sampling is provided to include the survey instrument and analysis methods used to identify relevant experts for each query and to also exploit survey "voting" patterns so as to classify experts as "authorities", "brokers" or both. Overall, these data and process descriptions promote methodology transfer to other settings allowing comparison to the results obtained here.

In performing the actual experiments, experimental data archival was limited primarily due to corporate policy which precluded long term retention of selected metadata used in expertise ratings and raw evidence in the form of relevant artifacts, and activities. While this was not an obstacle in running the actual experiments over a few days, it is problematic in terms of rerunning experiments at a significantly later date with the same collection; say, to study the effect of parameter changes, or the impact of alternative methods. Essentially, changes in the underlying environment introduce new evidence sources, confounding direct comparison of system performance between current and future experiments. For example, every month there are on average 300 new postings per ListServ; or approximately 60,000 postings per month across the 2000 ListServs analyzed in the evaluation. A one year lag between experimental runs would find roughly 720,000 new ListServ posts.   Overall, the following data were archived.

- Test Queries (29)
- The relevance baseline (qrels) in the form of snowball sampling results for each query, to include raw survey data, snowball experts lists, and associated hub and authorities scores for each survey mention.
- System settings used for each experimental run; to include the methods used to assign activity space weights, artifact/social evidence weights, retrieval depth parameter, and the total retrieval list size (e.g., retrieval limit is 100.)

- System output includes the *Expert Locator* ranked retrieval list for each test query to include selected organizational attributes

1.4   Thesis Organization

The remainder of this thesis describes the research underlying the development, deployment, and evaluation of the *Expertise Locator* prototype.   The work is described in the following chapters:

- **Chapter 2: Expert Finders**: this chapter introduces *expert finders*; tools that create awareness by cross-cutting knowledge silos to broaden perspectives on organizational expertise. Historically, expertise awareness has been addressed largely as a "retrieval" problem in which the goal is to identify expertise indirectly through lower-level retrieval operations that match work artifacts to expertise queries. Awareness is viewed here as a type of *finding* operation applicable to manual search strategies across personal relationships as well as automated methods developed largely in the database and information retrieval (IR) communities. However, while expertise search engines are of central interest here, the focus is broadened to include implicit "finding" operations embedded within organizational workflow and community services. This follows the notion of "ambient findability", Morville (2005), where information location or access is viewed from the perspective of being embedded within a particular work context or surrounding. As such, *finding* is not strictly aligned with the query-answer paradigm, but suggests a wider range of methods that make expertise *locatable.*

- **Chapter 3: Expertise Signaling**: There is an extensive literature focused on *experts*, their characteristics and behaviors. Generally, experts are viewed as high-performers having superior knowledge and problem solving skills when compared to novices; however, this runs counter to what is known about expert's performance in various decision contexts where cognitive biases may contribute to poor performance in decision making or predictive tasks. Yet from this disparity emerges a constant: <u>experts signal their expertise</u>. Experts signal their skills and experience to advertise capabilities, build reputation, and establish trust. Signaling behavior is visible and provides a basis for detecting experts, identifying relevant organizational context, and mitigating the problem

of explicit expertise encoding. This chapter explores the nature of expertise, and lays the groundwork for the signaling-based expertise model presented in this thesis.

- **Chapter 4: Activity Space Model**: Chapter 3 outlined the basic motivation for an expertise search capability based on the notion of expert signaling. The underling premise is that experts signal their qualifications through specific *activities* and artifacts within some organizational setting. As such, the central unit of analysis is the *activity space* (AS); a sampling frame of sorts that binds expert signaling behavior to a particular work context. This chapter lays out key elements of the AS framework.

- **Chapter 5: Enterprise Activity Spaces**: the activity space model presented in Chapter 4 provides a *template* for identifying specific activity spaces in the MITRE environment and assigning them into categories. Using this model, a number of MITRE activity spaces are described here (and Appendix C) from the perspective of their use in the *Expert Locator* system. Integration of specific activity spaces into the expertise model and operational prototype is discussed in Chapter 6 and Chapter 7 respectively.

- **Chapter 6: Formal Expertise Model**: much of the expertise modeling literature is domain specific, and emphasizes use of domain knowledge and methods as discriminators between experts and non-experts Chi, Glaser, and Farr (1988). However, domain-specific expertise models are not easily generalized and applied to expert finding. To address this, the expertise model developed here associates expertise with expert signaling behavior: communication used to convey specific knowledge or expertise. The model is extensible so that signaling behavior from multiple activity space contexts can be fused into an aggregate retrieval score assigned to candidate experts. This expertise rating is used to rank experts.

- **Chapter 7: *Expert Locator* Prototype**: this chapter describes the *Expert Locator* system architecture, user interface, and functionality. Specific emphasis is given to systems engineering issues and design tradeoffs central to deploying the prototype into an operational environment while still maintaining design flexibility needed to support this research.

- **Chapter 8: Evaluation Issues**: in one sense the enterprise is a "hostile" environment in which to conduct an evaluation; there is a lack of experimental control compounded by operational constraints imposed by the host organization. Here, there was no existing

system to compare *Expert Locator* to, no training data to baseline the new system against, and no a priori knowledge of what constituted relevance for a given topic—inhibiting the development of a test collection. This chapter discusses how operational constraints factored into a number of key evaluation issues to include: test query generation, relevance assessments, and results scoring. While the evaluation model used borrows from large-scale evaluations like TREC, the evaluation of *expertise* relevance as opposed to *document* relevance required a new approach to building a test collection and to assigning relevance to <u>people</u> and not <u>documents</u>.

- **Chapter 9: Methodology and Results:** this chapter covers experiments used to assess *Expert Locator* performance to include measures of system robustness to variation in queries and sources of evidence used. The chapter begins by developing a survey-based relevance set generator using snowball sampling; the method produces consensus-based query-relevant lists for a number of expertise topic areas. This process sets the stage for the precision-based assessments that follow. The chapter also includes a discussion on alternative evaluation methods; in particular novelty measurements as a basis for assessing the amount of "new" information provided in retrieval.

- **Chapter 10: Conclusions and Future Work**: the final chapter reviews main findings and presents several areas for future work.

- **References**: This thesis is multidisciplinary as reflected in research citations covering relevant prior work in information retrieval, cognitive science, signaling theory, and activity theory. Chapter citations for which there is clear authorship are cited in the Reference section; works without clear authorship are cited generally in footnotes.

- **Appendix A:** *Expertise Locator* Survey Form**: t**he online survey form used to generate baseline relevance assessments is presented here.

- **Appendix B: Selected Precision Results:** precision results are given in greater detail.

- **Appendix C: Additional Activity Space Descriptions**: Activity Space definitions, from Chapter 5, are expanded here to include supporting statistics regarding evidence distribution, membership, and general usage where available.

2 Expert Finders

Expertise awareness is becoming increasingly important in large, complex organizations forming a basis for "knowing which users should be made aware of which other users, how should users be made aware of one another, and how should these users interact", Maglio et al (1999). However, as organizations become more diverse and geographically distributed, work complexity increases so that expertise is often compartmentalized; restricted to business or geographically-based "silos" that support vertical knowledge integration but lack cross-boundary connections to related work and supporting organizations.   This is exacerbated by privacy and need-to-know restrictions that limit information sharing and access to experts. As a result, in many organizations, awareness is mitigated by limited transparency of employees' knowledge and expertise.   Expertise awareness has social implications in that identifying *who knows what* suggests an integrated view of actors, work groups, and communities in an organization-wide social collective, Won and Pipek (2003).   This integrated view subsumes individual awareness, often framed in terms of help seeking or collaboration, as well as strategic awareness focused on work performance and collaboration across work groups, and communities-of-practice, Schlichter (1998).

This chapter focuses on *expert finders*, tools that create awareness; cross-cutting knowledge silos to broaden perspectives on organizational expertise. Historically, expertise awareness has been addressed largely as a "retrieval" problem in which the goal is to identify expertise indirectly through lower-level retrieval operations that match work artifacts to expertise queries. Awareness is viewed here as a type of *finding* operation applicable to manual search strategies across personal relationships as well as automated methods built around search tools developed largely in the database and information retrieval (IR) communities. However, while expertise search engines are of central interest here, the focus is broadened to include implicit "finding" operations embedded within organizational workflow and community services. This follows the notion of "ambient findability", Morville (2005), where information location or access is viewed from the perspective of being embedded within a particular work context or surroundings. As

such, *finding* is not strictly aligned with the query-answer paradigm used in traditional IR, but suggests a wider range of methods that make expertise *locatable*; making expertise awareness integral to everyday work practice and positioning expert finders as organizational workflow and problem solving enablers.

2.1   Organizational Perspectives

With the advent of corporate Intranets, and ubiquitous "sensors" to track work, expert finding is becoming a knowledge management (KM) enabler, Reichling and Veith (2005). However, while much of the original KM work centered on exploiting artifacts and large information repositories, expert finding shifts the emphasis from documents to people and activities. The distinction between expertise as "artifacts" and expertise as "social interaction" is addressed indirectly by Ackerman and Halverson (2003), who identified four technical directions in which to address expertise finding.   They single out *repository*, *expertise locator*, *computer-mediated place*, and *ad-hoc groups* as implementation strategies for making knowledge accessible and creating expertise awareness.   These viewpoints suggest an *access continuum* in which expertise finding ranges from "objectified" knowledge embedded in online collections, to tools and environments that mediate expertise exchange.

From an enterprise perspective, this suggests a design space in which expertise sharing mechanisms may operate autonomously or in some integrated fashion so that Ackerman's and Halverson's technical directions may not dictate orthogonal functionality but, instead, be viewed as interrelated design elements used to construct hybrid systems. This may range from *expert locators* implemented as social brokers connecting people based on expertise needs, to models that make *finding* implicit within virtual work spaces. For example, ListServs may be viewed as instances of *computer mediated environments* which enable expertise exchange through self-organization (*ad hoc groups*) around specific themes. Here, posted messages serve as "attractors" around which List members group and, depending on topic scope; multiple forums may synchronize to address a particular problem or information need. This can be augmented by notification services used to alert experts or others as to emerging topics and increase awareness as to *who knows what*.

All of this suggests that expert finding is situated within a potentially complex work environment so that multiple expertise organization and transfer mechanisms may be needed to increase organizational awareness of experts.   This includes methods that are adaptive and go beyond matching a priori specified queries to, instead, identify latent expertise dynamically, without user action. This is exemplified by the emergence of *active* expert finders used to support long-term resource allocation, or to provide a type of "just in time" delivery of needed expertise. Active systems provide users automated referrals to relevant work and experts without users having to ask for it. For example, the Human Knowledge Navigator[5] generates, as a background process, user profiles from observed work activities and related artifacts as the basis for dynamically matching users to activities. The system can be used across a range of applications; for example, to automatically customize learning contexts based on past e-learning sessions and related work activities; to populate meetings with participants that meet certain expertise needs; or provide help on specific problems or questions. While Human Knowledge Navigator dynamically maps expertise to relevant work contexts, Won and Pipek (2003) focus on making competencies transparent. They discuss a system, eXact, which works as a *notification-based* awareness system used to make visible individual or group expertise consistent with user work requirements. A three-level model addresses *work capture*, *expertise indicator extraction*, and *expertise models* referred to as "specificators". Essentially, expertise indicators are extracted from events associated with various sources. There is an event hierarchy in which simple events, (e.g. ListServ postings) may be combined to form complex event indicators. The system, while potentially complex, has inherent flexibility in terms of supporting a potentially wide range of expertise models which can be used to combine indicators as the basis for assigning expertise to a particular actor. For example, indicator A: "user X is a key member of ListServ Y" may be combined with indicator B: "user X has posted frequently on topic T" to ascertain expertise related to domain "D".

Systems like Human Knowledge Navigator and eXact suggest a multi-layer architecture in which expertise detection operates as middleware used to instantiate some work function. This is a

---

[5]   http://jp.fujitsu.com/group/labs/downloads/en/techinfo/technote/okar/knowwho-catalog200307en.pdf   The Advanced "KnowWho" with Semantic Web Technology: Human Knowledge Navigator. Accessed July 24, 2006

potentially useful framework in that it supports functionality such as *notification*, *negotiation*, *privacy* and *policy* adjudication, and other expertise sharing enablers. For example, notification in the form of user alerts may signal expertise relevant to user's work context, or it may work at a lower level to facilitate expertise exchange between agents or processes. Notification is a key element in operational environments where users lack awareness and where user initiated interaction is not practical. This is especially true where user workload, task priorities, and privacy govern notification protocol. This has been addressed in the eXact system where privacy, organization, and user filters are implemented so as to ensure expertise is captured and presented consistent with corporate policy, work practice, and user relevance needs. Interestingly, this allows expertise ratings to be adjusted consistent with personal definitions of what constitutes an expert and not only what a system may decide, say, based on statistical criteria. An application of this may arise in a research environment where a more experienced researcher might put less weight on a lower-tier conference paper as evidence of expertise, than a less experienced researcher or manager might.   Jokinen and Kanto (2004) used a similar strategy to adjust the response of a speech-based E-mail system based on user expertise assessments.   An adaptive expertise model calibrated users on several levels (e.g., user-system interaction) as a basis for increasing dialogue effectiveness.

Various architectures may support expert finding operations; design optimization depends on the operational environment and culture. Where centralized or broker-based models may be effective in one context, peer-to-peer models may have greater advantages in another. For example, SHOCK, System for Social Harvesting of Community Knowledge,   Lukose, Adar, and Sengupta (2003) provides a peer-to-peer framework for knowledge (expertise) exchange that provides privacy-protecting capabilities to anonymize user's web browsing or email activity. Essentially SHOCK clients can build user profiles that assess message relevance as a basis for presenting information to a user.   This model may have advantages where it is more effective to manage user personal information locally, at the client and under user's control.   This architecture may also be useful for supporting ad hoc groups or enclaves through targeted messaging or "channels".

While expert finders have been viewed as components of various information sharing architectures, a number of researchers have viewed expert finding as part of a larger problem solving framework. McDonald and Ackerman (1998) identified 2 phases: *identification* and *selection*. *Identification* involves search operations used to discriminate experts from non-experts while *selection* has more to do with which experts best satisfy the expertise need; for example, to support a task. Most expert finders focus on expertise *identification* and, much like current document retrieval systems, relevance ranking provides a default basis for *selection*. However, as in document retrieval, ranking may not necessarily align with user selection criteria. In particular, ranks may not mirror user needs with regard to expert's availability, physical location, organizational role, and current tasking; quite often the top ranked expert is not the one most suited for a particular task when all factors are considered. Viewing expert finding as consisting of one or more elements of a larger process serves several purposes. First, it decomposes expert finding into multiple components such as, query formulation, identification, and selection. Second it ties expert finding to an end-to-end problem solving framework that contextualizes lower-level expert finding operations. Expert finding as a problem solving component has technical aspects that drive implementation architecture, but alternatively, provide insights into qualitative, social views essential to understanding the role that expert finders play in actual work settings. This is reinforced in a number of work domains; for example, the everyday work of service repairmen.

Expert finding as an element of human problem solving is well depicted by the plight of Xerox repairmen, Orr (1996). Here, expertise finding amongst service repair specialists is largely supported by informal information sharing; that is, *telling stories*. Technicians talk about machines and their idiosyncrasies through an informal knowledge network and knowledge is transferred through stories as well as written service reports. More often, however, technicians find answers to tough problems by largely consulting with other technicians, and their daily work is organized so as to facilitate these informal information exchanges. In this environment, formal documentation and organizational communication, valuable for common repairs and initial training, are less critical to finding highly specialized expertise and "stories" of machine idiosyncrasies.

Service technicians working together to solve problems through expertise sharing is an instance of collective problem solving where cooperation is used to provide effective solutions to complex problems, Clearwater et al, (1991). Here, collective problem solving consists of repairmen communicating "hints" to other repairmen with varying expertise; each often providing partial solutions to an overall problem. Collective problem solving and notions of coordinated work (i.e., workflow) suggest the need for a larger integration framework. However, while a problem solving framework especially suited for expert finding is largely lacking, it is reasonable to use as a starting point the MacDonald-Ackerman two-stage model augmented by the problem stages developed by Wooldridge and Jennings (1994). Combining the two models produces the following problem solving framework:

1   **Problem Recognition or *Need***: A user (read: *user or agent*) recognizes the need for expertise, say, as a basis for obtaining help or for collaboration.

2   **Query/Needs Formulation**: A user translates an expertise need into an expertise needs statement or more specifically an expertise signature consistent with a particular search strategy.

3   **Expertise Identification**: An expertise needs signature is matched against expertise profiles. Based on an expertise model, candidate experts are ranked according to expertise level or potentially other "state" criteria; such as actor's availability for tasking and this may support operations such as user *selection*. While this phase suggests a query-answer paradigm in which expertise profiles are used to query some sort of collection; other models may be supported such as peer-to-peer or self-organization. For example, using the ListServ case discussed earlier, expertise models may be based in part on discussion thread characteristics used to identify "key persons" such as those having certain expertise based on their discussion role and information exchange. Discussion threads are organizing mechanisms built up around a self-organizing theme.

4   **Expertise Selection**: Candidate experts are selected based on expertise ranking or task-related criteria such as availability, experience level, or resource costs.

5   **Plan Formation**: Selected experts are aligned with problem solving activities; that is, selected experts are mapped into specific roles or task assignments.

6   **Task Activity**: Work is performed; experts apply knowledge and skills to the task.

7   **Monitoring/Feedback**: Work performance is monitored and assessed. Performance measures provide feedback to earlier stages as a basis for improving expertise profiling, search performance, or other operations.

Stages 2 and 3 align with typical expert finding scenarios in which system performance is viewed largely in the context of stated expertise needs and some basis for adjudicating experts from non-experts. This is similar to methods described in formal evaluations such as TRECENT 2005[6], Craswell et al (2005), in which an expertise needs statement is matched against expertise indicators as a basis for ranking experts. Qrels (query relevant lists) are used to assess system relevance across a range of queries. However, currently, there are limits to the extent that TRECENT and similar evaluations can provide a rich organizational context in which to frame an evaluation. In particular, privacy constraints restrict access to richer organizational work context that could be used to build more robust search models or to support more task-specific evaluation. Current datasets are limited and preclude capturing complex work flows, organizational structure, and cultural aspects used to address actual (operational) selection criteria. This necessarily reduces TRECENT emphasis on *selection*, i.e., Stage 4, where the focus is more on task and organizational context used to support selection and work assignment. In actual operational settings, selection is situated and conditional on the needs of the expertise consumer, task characteristics, corporate culture, and various "state" variables to include organizational assignment, availability, and location.

---

[6] More explicitly, to the Expert Search Task

Expert finding is purposeful and often focused on addressing skill needs within the context of a particular task. As such, Stages 5 and 6 are associated with work planning where required expertise is coupled with characteristics of the work assignment or overall resource need. This is exemplified in expert team building described in a futuristic NASA collaboration scenario, Becerra-Hernandez (2000):

> You are working in a project to build a new cryogenic handling storage facility. You encounter a problem, where upon testing, a valve fails. There is a design problem. You have two choices:
>
> - The first choice is to go back through the same process with the same company and NASA engineers working the problem
> - The second choice is to use Expert Seeker to organize the Rapid Answer Collaborative Knowledge Expert Team (RACKET).
>
> Using the expertise keyword 'cryogenics' Expert Seeker finds the following experts:
>
> - A collection of scientists from the University of Arizona for cryogenics studies;
> - A valve manufacturing expert from a plant in Detroit;
> - A cryogenic expert that worked on problems during shuttle that transferred to Marshall Space Flight Center.
>
> In addition, the Expert Seeker uncovers a collection of technical white papers and lessons learned that NASA has published from similar projects. The RACKET collaborates by video teleconference and the Internet to pinpoint the design problem, identify a feasible solution, and fixes the design problem in two days.

However, while the notional Expert Seeker tool described above suggests powerful expert finding and context analysis capabilities; current automated systems can not easily incorporate planning and assignment knowledge in support of query formulation, search, and selection. As such, team building and task assignment remain largely as extensions to core *finding* services. Finally, Stage 7 addresses experts' performance with regard to a particular task; essentially, performance metrics are used to assess the appropriateness of task assignments as well as the

accuracy of the expertise ratings used to support the selection operations.   For the most part, expert finding systems (research and commercial) do not support expertise ratings based on job performance or work relationships[7].

The model discussed above suggests a pipelined workflow; where model stages proceed sequentially with simple serial dependency. However, this may be misleading; especially when the model is applied in operational environments. This is addressed at a high level in Figure 2-1, where the seven stages have been organized into four aggregate stages for simplification: *Need, Find, Exploit*, and *Evaluate*. This chunking of the lower-level stages allows for simplification of what may be fairly complex feedback loops used to adapt the overall expert finding process. In particular, here, there are two main feedback paths. The *Assignment Feedback* path uses performance information to adapt ongoing task assignments, or to assess expertise gaps. In principle, expert-task assignment mismatches could be used to adjust upstream retrieval operations. This suggests the need for *Query-Retrieval and Assignment Feedback* loops that use performance data to adjust query/profile generation, weighting schemes, and selection criteria. Here, performance feedback is task-specific so that it can address domain-specific needs of high-precision retrieval environments. It also allows for user models to be used to adjust performance assessments consistent with the consumer's knowledge and performance criteria within a particular domain similar to that provided by the eXact system.
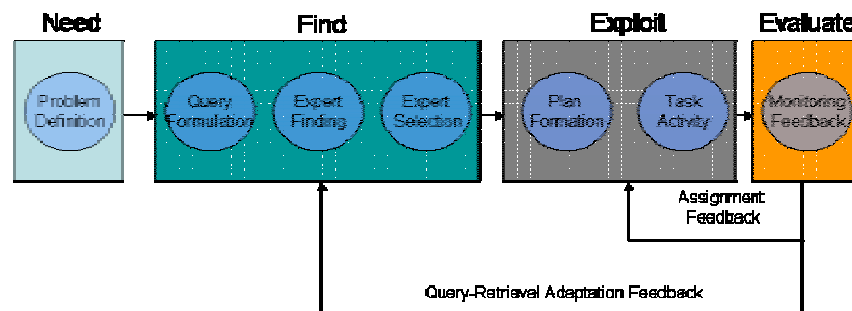
Figure 2-1 : Expert Finding Framework

---

[7] As noted by Resnick et al (2000) there are problems with eliciting, distributing, and aggregating performance feedback. Often people are not inclined to provide feedback, or may not provide a balance of negative and positive ratings. For example, registered "complaints" or negative opinions regarding another's performance may have long term implications regarding formal performance reviews or future work relationships. In other cases reviewers may lack skills necessary to provide accurate feedback.

Expertise finding has been viewed here in the context of overall workflow. This provides a potentially rich context in which to view the end-to-end expert search process; one that effectively couples simple finding operations to expertise usage and performance feedback. While there is simplification in some of the adaptive feedback loops it is robust as to underlying retrieval architectures; supporting centralized and distributed search architectures, and implicit and explicit finding operations.

2.2   Social Aspects of Expert Finding

Xerox repairmen, *talking about machines,* create collective expertise through bottom-up knowledge exchange absent higher-level directives to shape interaction. The process generalizes to a wide range of enterprise settings and is inherently social as workers identify experts based on referrals, search operations, or prior knowledge of *who knows what*. Trust amongst co-workers is built up over time and based on consensus of expertise and reliability ascribed to peers. Local knowledge of skills and experience, in the aggregate, leads to *organizational* expertise reflecting a collective view on actors, their roles, skills, and work relationships, Leibowitz (2001). As such, expert finding writ large, views expertise as widely distributed and not restricted to only a few individuals Huber, (1999). For example, open communities, such as online investment discussion boards[8], can have hundreds of participants with widely varying investment skills and considerable variation in knowledge of companies or industry sectors. For a given investment question, finding a single expert source may be insufficient; since in a wider context, expert opinion may vary considerably, on, say, what a fair trade price is or whether recent news suggests reduced profit, a reverse stock split, or delisting from a major exchange. Here, the trader may need to find expert investors on several boards to include expert opinion reflected in analyst's reports, and company news in order to assess a particular investment strategy.

Finding informed opinions from expertise "collectives", is problematic in large, complex organizations where expertise is obscured by rapidly changing work; geographically dispersed

---

[8] For example, the Google Inc. discussion board is found at http://messages.finance.yahoo.com/mb/GOOG; Accessed on October 10, 2006.

workforce; and cultural constraints. Reduced expertise visibility can impact organizational effectiveness in facilitating new employee integration[9], knowledge sharing and collaboration, Dixon (2000), and in mitigating the effects of lost expertise through workforce aging, De Long (2002).   The impact is wide-ranging and supports the need for strategic views on expertise; perspectives that span multiple domains, work settings, and diverse cultures. This motivates a social perspective; viewing experts as embedded within a rich socio-cultural context; an expertise network.

Expertise networks can be defined as "… specializations of an organisation's social network. They consider not only how people are socially arranged but what expertise they have and trade," Ackerman et al. (1999).   While organizational structure is important in identifying connections between work domains, expert finding has been largely a bottom-up process, centered traditionally on an actor's personal network and formal authentication such as academic or professional ratings and honorifics. As such, expert finding often involved exploiting personal contacts either through face-to-face contact, by phone, or through intermediaries. Li et al (2006) viewed personal contacts from a social network perspective in which, an arbitrary network node, Figure 2-2, searches the social network for nodes satisfying some query or expertise description. In the example network, person nodes are connected by four types of relations (*knows, collaborates, collaborated*[10], and *consulted by*). The relations characterize the association between nodes in the sample network, and whether there is reciprocity or not. With that, expert finding is framed as a graph search problem where the search space is constructed around social network members and their relationships.

The implication here is that from a graph traversal perspective, searches must align with graph topology; that is, search is constrained by graph structure.   The actual search strategy may be complex; for example, beyond the "simple" case of assessing nodes that the search node is directly tied to, other instances require resolving tradeoffs between shortest paths and utility. In effect, the shortest path between a starting node and a "target" node may not necessarily yield the

---

[9] "Mellon Learning Curve Research Study" discusses the problems of rapid *on-boarding*. www.Mellon.com; Accessed on August 23 2003

[10] Here, there is a temporal distinction between *collaborates* (i.e., a current activity) and *collaborated* (i.e., historical association).

most reliable or accurate information. For example, in Figure 2-2, consider a case where node (a), someone seeking an expert in temporal analysis, obtains information on the target node, (g), following several paths (i.e., referrals). Path (a)→(h)→(g) is "short" in that it provides evidence of (g) through node (h); a single hop. However, while node (h) is valuable in ascertaining expertise of (g); node (a) only *knows* (h); that is, (a) may not have a reliable basis for trusting (h). This is contrasted with the search path (a)→(e)→(f)→(g). This path has two intermediaries but each tie connecting (a) to the target (g) is associated with actual *collaboration*; in addition, the relation (f)→(g) refers to *current collaboration*. As such, in this hypothetical search graph, (a) must weight the value of evidence gained through a short path involving weaker ties compared to a longer path that is based on stronger (possibly more reliable) linkages. This suggests that the weight of importance placed on retrieved evidence is a function of value ascribed to node attributes and social relations.

Graph structure also has implications for search coverage; for example, in a directed graph, some nodes may not be reachable due to one-way relationships. Therefore, for a given query, the social network may be weighted as to "reachability" or to the utility of various nodes as to their query relevance or use in brokering ties to true experts. This local view (situated within the context of a query) is juxtaposed to the global view of nodes. Some nodes are inherently more "valuable" in supporting a local search; while other nodes have collective value based on their network position allowing for broader views on "who knows what". For example, node (e) is more central than node (b) in terms of connections to others and has more substantive ties to neighbors.
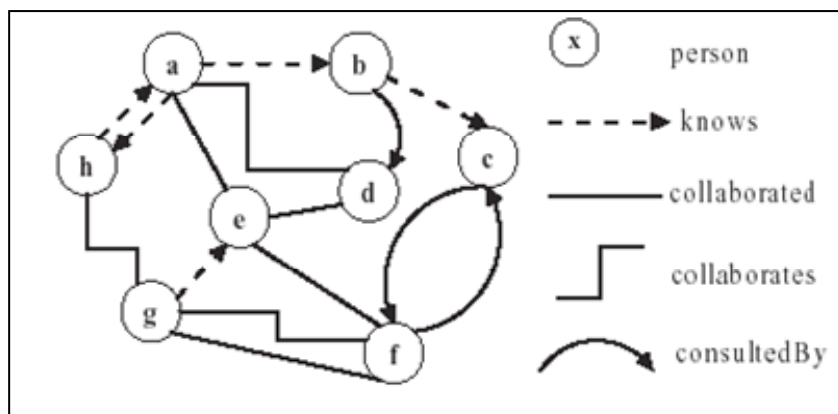
Figure 2-2: Social Network Schematic adapted from Li et al (2006)

Embedding experts within an organizational context recasts expert retrieval as a type of graph search problem where relevant subgraphs capture relationships between experts, artifacts, and social context. This extends traditional relevance assessments based on expert's attributes to also include embeddedness within relevant subgraphs representing work settings and links to others. This approach has been taken by D'Amore (2004) and earlier in the XperNet system, developed by D'Amore as described in Maybury, D'Amore, and House (2000), which extracted affinity graphs from larger social networks based on thematic overlap, co-work, and organizational structure. A commercial product, Parity's Profiler System[11], follows this model somewhat; it provides personal or organizational profiles that include network relationships to other individuals or groups. However, it is not clear to what extent it exploits expertise network structure in rating experts.

An expertise network is shown in Figure 2-3[12]. Graph nodes represent individual experts, and links between experts are based on co-work within a query-relevant topic area. Here the topic is *Biocomputing* and expert nodes are sized according to centrality; viewed here as a measure of importance in terms of an expert's connectedness to other experts.
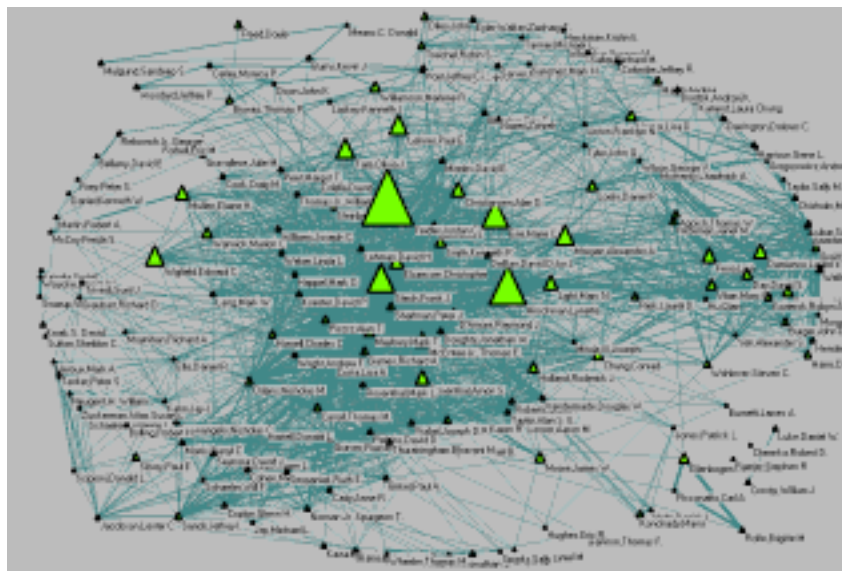
Figure 2-3: Expert Network Generated Using *Expert Locator*

2.3 Expert Finding System Issues

Direct application of traditional information retrieval and database search techniques, while effective for certain KM applications, is problematic for expert finding given the tacit nature of expertise. Unlike factual knowledge, expertise is not easily encoded, communicated, or shared. As such, the design, implementation, and evaluation of expert finder systems must address a range of issues as suggested by Pipek, Hinrichs, and Wulf (2003) and others:

- *Most critical knowledge is never made explicit in materials that can be electronically accessed.* Expertise is often obscured and not easily captured, processed, or transferred within organizations. Notably, there are a number of cognitive limitations related to problematic nature of tacit knowledge elicitation from experts, Epple, Argote, and Murphy (1996); expertise sharing across skill levels, Finkel, Heath, and Dent (2001); and cross-domain knowledge transfer, Langer and Imber (1979) and Hansen (1999). Underscoring this is the notion that expertise is "compiled" information not easily decomposed into chunks for easy encoding and reuse, Du Boulay and Ross (1991). This precludes easy capture and transfer through documents, presentations, and other artifacts which both reduces work efficiency and obscures expertise, Hinds (1999).

- *Data may exist in electronic form but be inaccessible for practical purposes because it was catalogued according to a system that has no relation to potential needs for that information.* This is consistent with the notion of expert's use of specialized terms and concepts not easily transferable to non-experts or across domains. This is also an instance of the decontextualization/recontextualization problem; noted by Ackerman and Halverson (2004). In order to reuse information (i.e., transfer expertise), it may be necessary to remove context (*recontextualized*) in order to form shareable boundary objects.

- *Organizational culture and policy can constrain development and evaluation*: Expert finder systems must be synchronized to corporate policy and privacy constraints if they are to be effectively integrated into corporate workflow. This not only imposes limits on the kinds of information users can access or share, but it also constrains system use so

that it is aligned with business practice and processes. This is especially problematic in organizations where there are significant privacy constraints or where work sensitivity precludes sharing or provides disincentives to making work visible, Hinds and Pfeffer 2003. Further, in many organizations, experts are tightly embedded in project areas and can not provide expert consulting on an open basis. This ensures that experts are aligned with formal work but it may inhibit informal exchange of expertise and shift expert's motivation from one of sharing, to protecting competitiveness by shielding knowledge and skills from others, Davenport and Prusak (1998). Related to this, experts may perceive risk providing help or advice in areas where errors or miscommunication may be detrimental to their status or formal position within the enterprise. While there are instances where experts have formal (i.e., legal) protection, say, via peer review, Hall (2006), in many organizations expertise exchange is brokered informally.

- *Evaluation baselines are difficult to generate and maintain; especially in operational environments:* The nature of expertise, discussed more fully in Chapter 3, presents a mixed view of expert's capabilities; knowledgeable and efficient yet prone to miscalculation. More so, the decision analytic literature identifies a number of performance deficiencies related to analytic biases such as anchoring and availability that restrict their ability to assess alternative solutions or properly weight evidence, Tversky and Kahneman (1973). This has tactical implications regarding expert performance assessment on a particular task and in developing expertise rating schemes that are transferable across expertise domains. As such, expertise evaluation is inherently problematic and dependent on qualitative assessments or on quantitative measures limited to particular work contexts or narrowly framed tasks. This has significant implications both for the assessment of core finding algorithms as well as for the incorporation of enterprise context and peer ratings into overall expertise assessments.

Many studies of expert's performance are baselined on ground truth where there is some notion of an optimal or correct result. For example, the TREC Enterprise Track (TRECENT) developed a relevance baseline for email messages, Web and other extranet data collected from working groups at the W3C[13], Craswell, Zaragoza, and Robertson

---

[13] World Wide Web Consortium (W3C)

(2005). In an effort to shift the focus more to the expert search problem and less on collection construction, relevance assessment was simplified: topics were equated with groups and the experiment goal was to correctly retrieve people who were members of a particular group. In follow-on work, for the year 2006, a more fine-grained approach was taken where approximately 20 groups contributed to 55 topics; with roughly 2 to 3 topics selected from each of the submitting groups. Each group judged their own topics as well as topics from other groups; most groups were expected to judge approximately 6 topics. The results from multiple participants were pooled[14].

Document relevance is central to TRECENT expertise ratings; that is, relevant documents associated with a particular person form the basis for expertise ratings or rankings. Here, document relevance is a function of document-query similarity; while there is some allowance in practice for other factors to include user background (e.g., expertise), and search context, for example, the order in which documents or expertise evidence are viewed. However, operational expert finding systems may base expertise ratings on a wider range of evidence than simply documents. For example, an actor's level of participation in a data mining project may be used to assess expertise. In addition, document (i.e., artifact) relevance does not always convey expertise. For example, highly relevant documents may be discounted if they are associated with a work context that is not assessed as relevant to a particular work practice or organization. As such, an operational expertise model may introduce relevance criteria that go beyond those used in typical document retrieval environments. Limitations in using document relevance as a basis for assessing expertise does not diminish the emphasis placed on establishing some kind of expertise baseline in which to assess system performance. While knowing *who knows what* may be practical in certain experimental environments it may be impractical in operational test settings where experts may constitute an unknown population for certain topical domains.

---

[14] Much of the relevance assessment work is captured in the List discussions, http://cio.nist.gov/esd/emaildir/lists/trec-ent/threads.html Accessed on September 10, 2006.

The absence of a "correct answer" is commonplace in many real world environments so that measuring expertise when no "gold standard" or correct answer is available is problematic; Gigerenzer et al., (1999), and Shanteau et al (1993). Performance based measures (PBM) have been advocated; for example, Shanteau et al (2002). Early efforts to provide a PBM include the use of subject matter experts (SME's) where answers are obtained by simply querying an expert. However, expert's accuracy can vary considerably depending on the domain and the type of decision support tools used, Shanteau (1992). For example, weather forecasters have demonstrated better prediction performance than financial market forecasters.

Shanteau also points to use of two characteristics: internal discrimination and consistency. When an expert is internally <u>consistent</u>, then consistency may be associated with expertise when combined with measures of discrimination (the ability to make distinctions). Shanteau proposes the CWS[15] statistic which "is based on the idea that expert judgment involves discrimination – seeing fine gradations among the stimuli – and consistency – evaluating similar stimuli similarly", Shanteau, (1993). CWS, the ratio of discrimination to consistency, has, in some domains, been shown to have utility as a measure of expertise level; however, it has limited usefulness in areas where there is a weak basis for quantifying either of the two measures used in the ratio. The problematic nature of assigning relevance within a particular evaluation setting, complicated by the variation in metrics used across various studies, contributes to the problem of transferring results across domains and precludes easy comparison of competing methods. While TRECENT provides a useful framework for assessing multiple technologies within a controlled setting, it does not currently provide sufficient organizational context (i.e., a dynamic work setting), to assess the operational effectiveness of any particular system.

- *Missing Experts:* In large evaluation environments, it is difficult, on average, to identify all relevant experts. Establishing an expertise baseline of known relevant experts for a range of topics is not addressed well by random sampling or through centralized committees or panels. This is a central issue addressed in this thesis as discussed in Chapters 9 and 10. The basic issue here is that missing experts (here viewed as

---

[15] Cochran-Weiss-Shanteau (CWS)

unknowns) can skew the evaluation and must be addressed in a consistent manner. This
problem is endemic to IR evaluations as discussed in Buckley and Voorhees (2004).

2.4   Expert Finder Systems and Services

The section focuses on the use of expert finders within external communities and enterprise
environments.   This partitioning is useful in that it naturally groups tools and services that work
within formal intranet environments as one class and those that are associated typically with
"non-critical" computing environments in another.

2.4.1   Community-based Services

Expert finders are becoming increasingly common in online (virtual) communities; providing
users a way to identify special skills or to find individuals with common interests. Virtual
communities vary considerably with regard to focus and membership, and, while there is no
consensus on what constitutes a virtual community, numerous working definitions abound.
Virtual communities have been described as "…social aggregations that emerge from the Net
when enough people carry on those public discussions long enough, with sufficient human
feeling, to form webs of personal relationships in cyberspace", Rheingold (1993).   While this
definition emphasizes broader, communal aspects others have focused on some of the lower level
mechanisms necessary to support social interaction and information exchange. Whittaker et al
(1997) characterized communities based on the presence of core attributes such that
"communities with more such attributes were clearer examples of communities than those that
had fewer. " The identified attributes, below, were viewed as "indicators" of social organization
or cohesion where members interact based on common purpose; with information and
communication services guided by policies:

- members have some shared goal, interest, need, or activity that provides
  the primary reason for belonging to the community
- members engage in repeated active participation and there are often
  intense interactions, strong emotional ties and shared activities
  occurring between participants

- members have access to shared resources and there are policies for determining access to those resources
- reciprocity of information, support and services between members
- shared context (social conventions, language, protocols)

Similarly, Selznik (1996) identified seven elements of community: *history, identity, mutuality, plurality, autonomy, participation*, and *integration.* Here, there is emphasis on individual autonomy and collective action; however, more interestingly there is emphasis on community memory. This suggests viewing communities as learning organizations, Huber (1991), able to capture and retain knowledge over time and attribute it as to source and transfer mechanisms. Collectively, community characteristics as suggested by Selznik, Whittaker and others suggest a rich social and information context within which to embed expert finder services. They promote expert finder services that exploit not only member characteristics, but also social interaction and sharing mechanisms. The importance of capturing community history through a shared memory is suggested; providing a basis for tracking community expertise across members and activities.

The focus is narrowed further by segmenting communities into *communities of practice* (CoPs), Lindstaedt (1996), Lave and Wenger (1991), and *communities of interest* (CoIs), Fischer and Ostwald (2001). While in principle both CoPs and CoIs may exploit similar communication and information services, CoPs, as used here, have a single domain focus, generally, while CoIs are often multi-domain. For example, while AllExperts[16] is an open question-answer based CoI covering a wide range of topics, SeniorNet[17], is a CoP organized to "provide older adults education for and access to computer technologies to enhance their lives and enable them to share their knowledge and wisdom", Mynatt et al, (1999). Here, peer-to-peer interaction is mediated through various communication services such as email, chat, and ListServs.

Aside from communities, other network models may also be applicable. For example, a dynamic team-based organizational framework called "knotworking", Engestrom et al (1999), may have applicability in modeling dynamic team formation, for example, certain types of informal work,

---

[16] http://www.allexperts.com/ Accessed on August 23, 2006
[17] http://www.seniornet.org/php/default.php Accessed on 2 April, 2007.

where expertise is viewed from the perspective of rapidly formed and disbanded teams in which teams are not persistent and are driven by dynamic tasking. More specific to the *Expert Locator* model, personal networks, which are related to intensional networks, Nardi, Whittaker, and Schwarz (2002), provide an ego-centric view of expertise that shifts emphasis from experts as members of groups and larger organizations, to experts as central actors within a particular organizational neighborhood. Importantly, personal networks are an integral part of the *Expert Locator* prototype developed in this thesis. Selected expertise services supporting CoPs are discussed next.

2.4.1.1   *Community-of-Practice Based Services*

CoPs are associated typically with a particular domain and built around a central, organizing theme. For example, Lesser and Storck (2001) define a CoP as "a group of people playing in a field defined by the domain of skills and techniques over which the members of the group interact". Similarly, Fischer and Ostwald (2001) note that "Communities of practice consist of people sharing a common practice or domain of interest." They further emphasize that "CoPs are sustained over time" and "provide a means for newcomers to learn about the practice and for established members to share knowledge about their work and to collaborate on projects." A central focus here is the need for special support to ensure community members understand the "long-term evolution of artifacts and for understanding problems caused by rapid change in their domain". As such, the need for supporting infrastructure and shared principles necessitates a *common ground* be established, Clark (1992) and Clark et al (1983).

Common ground is addressed in a number of environments through the use of registration-based services that assign experts into pre-defined expertise areas (typically through some enabling taxonomy.) ProfNet Experts[18] provides journalists with access to experts who "can comment on newsworthy topics in daily ProfNet Wire feeds". The system supports a number of user types to include reporters, information officers serving as search intermediaries, and actual experts submitting expertise profiles. Several screen captures are shown in Figure 2-4. A key aspect of CoP systems like ProfNet Experts is the expertise profile management system. This system

---

[18] http://profnet1.prnewswire.com/login_prn.jsp Accessed on August 14, 2006.

serves to capture expertise through a registration process; experts assign into particular expertise categories and enter expertise descriptions as shown in the template on the right side of the figure. There is typically an adjudication process used to validate expertise but for the most part experts self-assess their skills and experience. There are various services that go beyond search to include a profile linking service that links expert mentions in news articles to stored profiles in the ProfNet Experts Database.

Figure 2-5 provides a view of Newswise; a system used to "distribute news to journalists who have requested it." Newswise also provides journalists with access to domain experts through directory services and automated searching. Journalists can search contact directories to find specific expertise; however, often the primary interface is an organization point-of-contact acting as an intermediary to actual experts. Users can also search for experts in past Newswise articles.
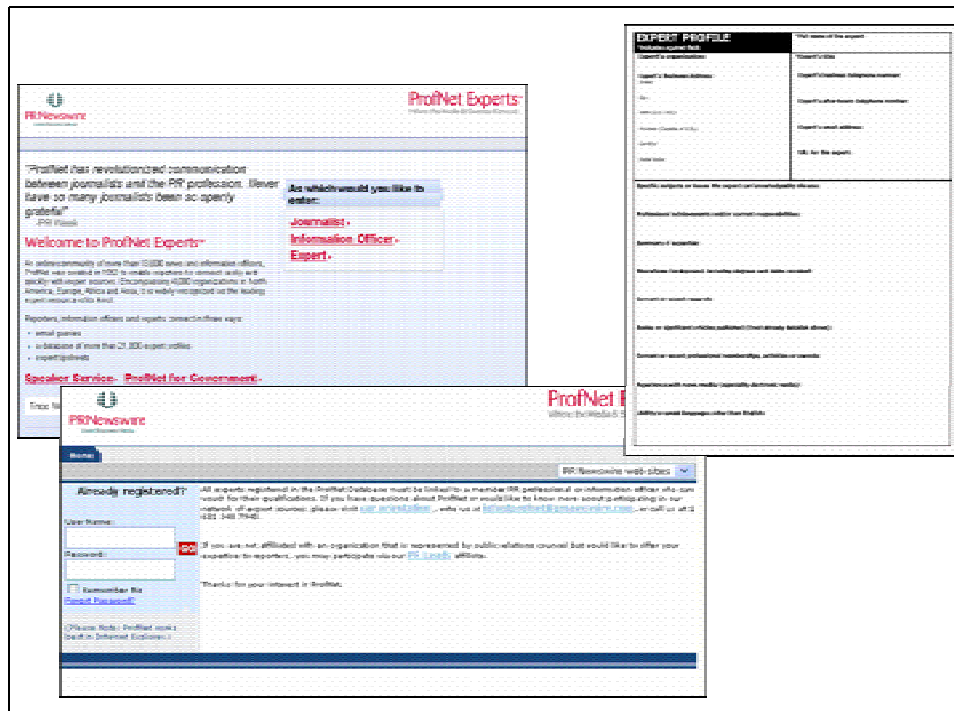
Figure 2-4: ProfNet Experts: A Typical CoP Expertise Registration System
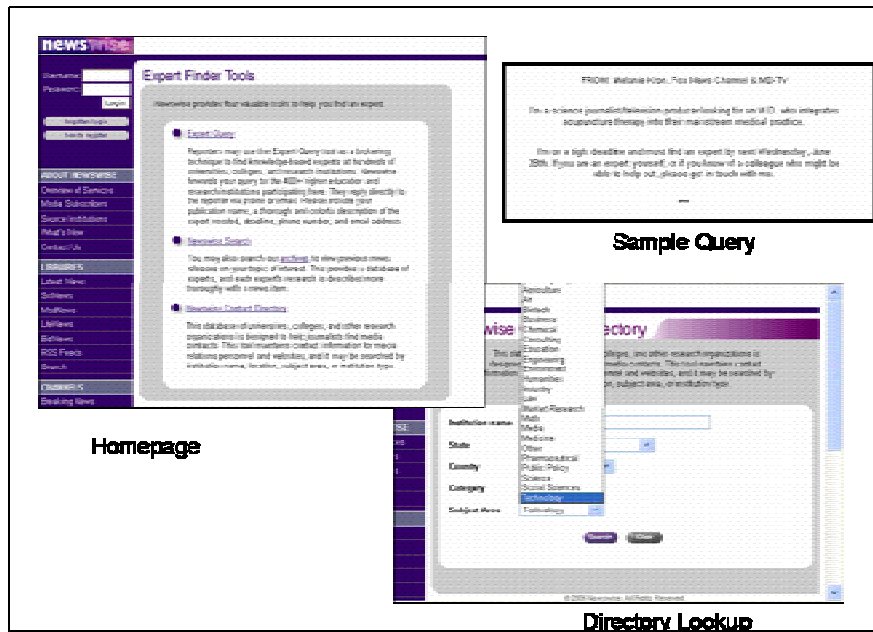
Figure 2-5: Newswise Experts: Directory services provide access to experts.

Additional expert finding services are listed in Table 2-1 and characterized with regard to built-in support for expertise representation, expertise ratings, and expert search/browse operations. Although most systems are not distinguished by any particular technology innovation, e.g., none employ advanced expertise models; collectively they suggest the kinds of extensible architectures needed for managing expertise within specialized domains that may cross-cut multiple organizations and diverse user populations.

Expertise Service	Expertise Directory	Expertise Topics	Search/Browse Capability	Expertise Profile	Expertise Adjudication
EIN[19] The Ecological Information Network (EIN) is a database of ecological experts who voluntarily answer questions or provide input on various scientific issues.	Expert listings organized by topic.	Eight main areas and numerous sub-areas.	Browse lists of experts by topic areas; no search capability	Limited: topic label, affiliations, and contact information	Registration process; adjudication unknown
WTB[20] (World Taxonomic	Directory services to	None provided	Search for persons, institutes,	Self-classification	Registration process;

---

[19] http://ein.nbii.gov/EIN/index.faces Accessed on August 18, 2006

[20] http://www.eti.uva.nl/tools/index.php Accessed on August 18, 2006

45

Database) ETI's World Taxonomist Database, an online directory service includes information taxonomists, specialists worldwide.	over 4000 scientists and specialists		country, or group.	by taxonomic group (order, family, genus), environment, geography,	adjudication unknown
Community of Science (COS) Expertise Database[21] is a knowledge management system for individuals and institutions, with more than 480,000 first-person profiles of researchers from over 1,600 institutions.	Directory of registered experts	Taxonomies (e.g., social sciences) used to support expertise profile development and user searching	Browsing expert directories or searches against expert	Expert profiles include name, position/title, location, publications, memberships, and keywords).	Registration process; adjudication unknown
ProfNet Experts[22] An online community of more than 13,000 news and information officers, ProfNet enables reporters to connect with expert sources. ProfNet has 4,000 organizations in North America, Europe, Africa and Asia.	None provided directly to users.	Open. Topics defined implicitly through expertise profiles.	User queries matched against expert profiles. Email-based dissemination system used to dynamically alert users as to experts relevant to standing query.	"Resume" format supporting free-text entry. Experts enter key skill or experience areas, professional achievement, research, foreign language skills, and contacts.	Peer review supported by information officers who may also sponsor specific experts and monitor their activity.
Newswise Expert Finder[23] provides tools to help journalists find an expert	Contact Directory provides access to organization points of contact acting as brokers to experts.	22 Fixed Categories oriented around 4 News publications: MedNews, SciNews, LifeNews, and BizNews.	Journalists can browse or search contact directories. Queries are manually reviewed. Users can query 50k articles to find experts mentioned.	Expert from various organizations submit expert profiles and contact information.	Newswise internal review.

Table 2-1: Selected Expert Finder Systems Supporting CoPs

---

[21] http://expertise.cos.com/ Accessed on August 18, 2006

[22] http://profnet3.prnewswire.com/enter/index.jsp Accessed on August 18, 2006

[23] http://www.newswise.com/resources/experts/ Accessed on August 22, 2006

2.4.1.2   Community-of-Interest Based Services

Communities of interest, Fischer and Ostwald (2001), are made up of people with different
backgrounds who organize around a particular issue to share information, or take part in some
activity upon which there is some shared view.   While CoPs comprise collective knowledge,
within a specialization, CoIs are potentially eclectic with members having diverse backgrounds,
interests, and skills.

CoIs may impose fewer constraints than CoPs on membership, information sharing, and tool use.
As such, they bring together actors from different communities and diverse cultures who may
self-organize around particular issues or events. However, diversity may make problematic
establishing common ground as discussed above. A summary of discriminating characteristics
used to distinguish between CoPs and CoIs is provided in Table 2-2, taken from Fischer (2000).

Dimensions	CoPs	CoIs
Nature of problems	Different tasks in the same domain	Common task across multiple domains
Knowledge development	Refinement of one knowledge system; new ideas coming from within the practice	Synthesis and mutual learning through the integration of multiple knowledge systems
Major objectives	Codified knowledge, domain coverage	Shared understanding, making all voices heard
Weaknesses	Group-think	Lack of a shared understanding
Strengths	Shared ontologies	Social creativity; diversity; making all voices heard
People	Beginners and experts; apprentices and masters	Stakeholders (owners of problems) from different domains
Learning	Legitimate peripheral participation	Informed participation

Table 2-2: Differentiating CoPs and CoIs

As discussed, many CoPs use registration-based expertise services as a coordination mechanism;
experts self-assign into areas of specialization providing users with a coherent, domain-specific
mapping to expertise areas and actual experts. This approach, however, may not scale well in
CoIs which have widely varying domains and membership.   However, there are communities of
interest that exploit registration for capturing expertise from multiple areas. For example,

Google Co-op[24] is an open community system that provides a number of services to support effective access to information. A publish-subscribe model, allows users to subscribe to a particular topic as a <u>provider</u> adding content or as <u>consumer</u> using stored content to augment searches. Searches are augmented using a type of co-operative searching framework in which user queries are effectively expanded using results sets derived from information sources (for example, Web pages) provided by registered experts or organizations. Community expertise ratings are assigned to contributors based on the number of subscribers and frequency of topic use. Figure 2-6, shows a typical expert profile indicating which "expertise areas" a particular expert is assigned into, and which users have linked to that expert.

Figure 2-6: Google Co-op "Expert" Profile Page

While Google Co-op makes experts "visible" through registration and community ratings, Google Answers[25] masks experts from those providing questions. In this model, users pose questions for a small fee (typically $0.50) and attach the price they are willing to pay for the answer. Experts choose questions based on how they match up with their own expertise as well as the fee they will receive from the questioner. In this case, questioners do not have direct access to available expertise other than through a type of "negotiation". Essentially, if experts do

---

[24] http://www.google.com/coop Accessed on August 14, 2006
[25] http://answers.google.com/answers/ Accessed on August 22, 2006

not "lock" a question (i.e., choose to answer it), users may be forced to raise the fee. Users can only view expertise through the quality of the answers to their own questions or through prior question-answer pairs from other users. Users can rate answers; this provides a way for Google to manage experts in terms of future use.

Beyond, Google Answers, a number of systems are built around a question-answer (QA) paradigm. For example, Abuzz's Beehive[26] provided an on-line community environment to support question-answer dialogues between users and registered "experts". Users could learn from other user's question-answer dialogues posted under specific topics such as *cooking*. Communities of experts are grouped in *web circles* that provide a domain-specific context for registering as an expert, for users to ask questions or initiate a group discussion. 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-answer pairs similar to their own question. If users did not find a related question they were referred to a category-assigned expert.   The emerging on-line commercial systems, for example AskMe Pro[27], track each expert's 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[28]) 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 appropriate experts.

More recently, a number of similar QA sites have been developed such as the Mad Scientist Network[29] which "fields questions in 26 different subjects, covering topics in astronomy, the

---

[26] Formerly accessible at: http://www.abuzz.com/ Not publicly available; bought by the New York Times
[27] http://www.alstrasoft.com/askme.htm Accessed on December 8, 2007.
[28] www.xpersite.com Accessed on August 22, 2005
[29] http://www.madsci.org/ Accessed on August 22, 2006

biological sciences, chemistry, computer science, earth sciences, engineering, and physics."
Similar to Google Answers, the site uses moderators to screen questions and answers for quality,
to answer questions that are related to prior QA pairs or through online searching, or to forward
questions to appropriate scientists as necessary. Moderators, accessing roughly 35000 QA pairs
and other resources, are able to handle a number of the questions. However, as necessary,
moderators search approximately 700 expertise profiles to find experts that match a question and
forward the question to the appropriate expert. Again, like Google Answers, users do not have
direct access to actual experts but they do have access to prior questions and answers.

More recently, Liu, Croft, and Koll (2005) explored expert finding within Wondir[30] , an open
community, question-answer service. As reflected in the screen shots from Wondir, Figure 2-7,
users can view recent questions via a question ticker tape, select a question to answer, pose new
questions, and scan a question bulletin board for questions and answers. Answers are rated so
that experts build up a quality score based on total answers provided and average rating. One of
the higher rated experts is represented in the lower screen shot.

---

[30] www.wondir.com accessed on June 17, 2006.

Figure 2-7: Wondir Screen Captures: Query, Question Board, and Expert "Profile"

Open community systems, like Wondir, raise a number of issues as to how to assess expertise and build trust between users and candidate experts. Liu et al ran experiments centered on 852,316 QA pairs extracted from a slightly larger collection. As noted, expertise is ascribed to a user simply by answering the question; this raises some issues regarding the relevance baseline used in the analysis. This is reinforced by perusing the Wondir site (from which the data were collected) where browsing through a number of question categories shows many poorly formed questions, and on average fewer than 2 persons answered a question as noted in the study. In particular, a qualitative assessment of a number of the most prolific experts revealed they were also highly rated, had rapid question-answer turnaround, and provided very short answers. Interestingly, question-answer pairs took on characteristics of a topic thread in which pairs resembled chat sessions but with added latency. This suggests that the system is being used in ways that may support communication but this usage may not be effective in capturing actual expertise. The fact that question content may be problematic in inferring expertise is not unique to Wondir as reflected in the most common question phrases and keywords from the Mad

Scientist Network, Figure 2-8. These phrases, viewed essentially as (fragments of) expertise queries are short, have significant variability, and have low frequency of occurrence. The keyword list suggests users use "conversational" style to generate queries as noted by high rank afforded function words such as *of, how*, and *the*.
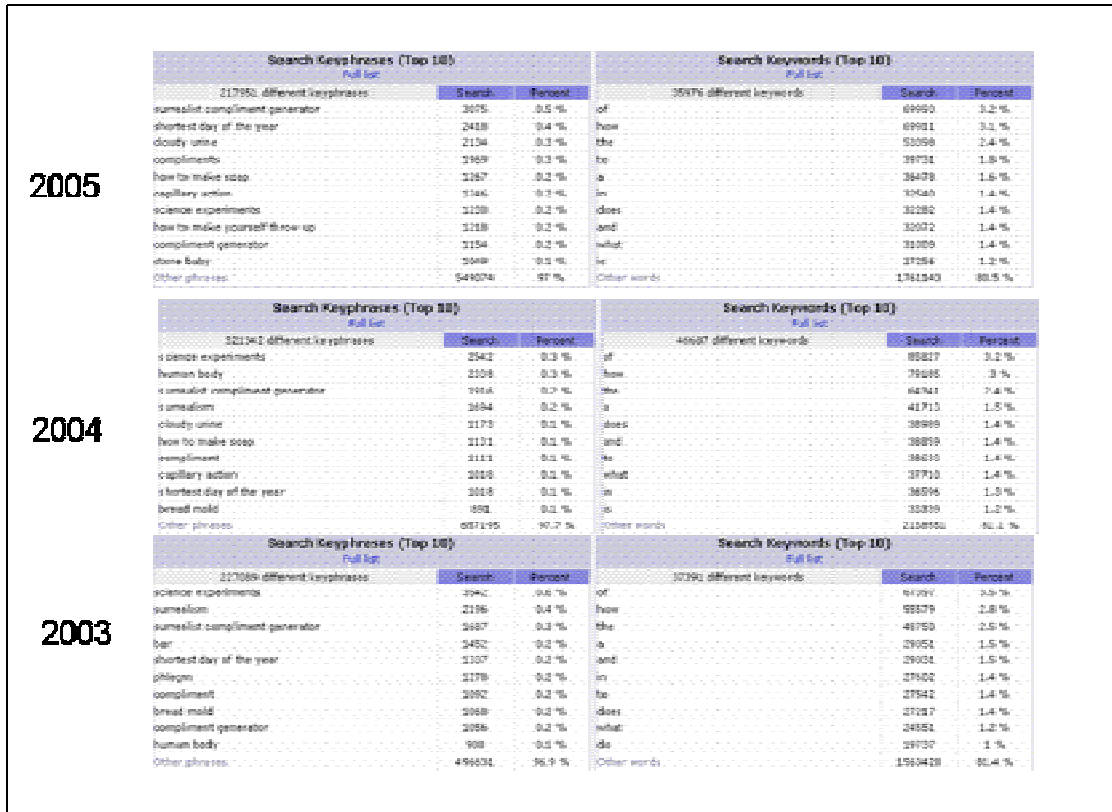
Figure 2-8: Most Common Question Phrases and Keywords over a three year period

Clearly, open communities pose significant challenges to expert finding as there is little constraint on the range of questions, little context that can be used to qualify expertise, and little visibility into how answers are formulated. For example, an "expert" could simply look up answers for certain question types which while the answer may prove useful, does not guarantee the answer provider is an actual expert. In addition, in many cases where questions have limited interest outside the questioner, there may be little incentive to provide alternative answers or to confirm the accuracy of answers already provided. Given the problematic basis for connecting actual experts to questioners, the system provides more traditional question-answer services as

well. For example, each question is linked to (Web) resources (e.g., web pages and documents) that may be relevant to the query. Table 2-3 overviews selected community-based QA systems that support expert finder services.

Expertise Service	Expert Directory	Expertise Topics	Search/Browse Capability	Expertise Profile	Profile Generation and Adjudication
Kasamba[31] "30,000 skilled professionals are ready to give you immediate advice or assistance on any conceivable topic…"	Expert listings organized by domain.	Eight main areas and numerous sub-areas.	Browse lists of experts by topic areas; no search capability	Varies across experts; includes area of expertise, education, affiliations, and contact information.	Self-declaration. Adjudication unknown. Community quality ratings associated with experts having a QA "history".
ExpertBee[32] "…has a community of experts that will bid to provide you with an answer. "	None provided.	Forty-four topics.	Post question within user-selected topics; experts bid on question. Winning bid establishes client-expert relationship.	"Resume" like profile.	Self-declaration; adjudication unknown. Consumers can provide feedback used to "rate" experts regarding quality, timeliness, etc.
Wondir[33] An open community QA system.	None Provided	Large number of topics	Browse/Search questions within selected topics. Search all topics.	Personal profiles not necessarily reflecting any particular expertise.. Anyone can ask a question or answer it.	Users have a short descriptive profile augmented by performance scores based on community feedback.
All Experts[34] "…a one-stop shopping source for free questions and answers on virtually *any* topic!"	Experts listings organized by domain	Thirty-six subject areas with subcategories in most	Browse list of experts organized by topic. Review expert profile.	Simple free-form text description.	Self-declaration. All Experts reviews application but as they note "you're almost certain to be accepted!"
Google Answers[35] is a way to get help from Researchers with expertise in	An expert directory is not directly available.	Archived QA pairs are organized into 8 groupings	Browse or search archived QA pairs. Expert attribution to each QA pair provides limited	Application process not available when site was last accessed, August 23, 2006.	Self-declaration through expertise descriptions and job postings. Experts are evaluated by

[31] http://www.kasamba.com/Default.aspx Accessed on August 18, 2006
[32] www.expertbee.com Accessed on August 23, 2006
[33] http://www.wondir.com/wondir/jsp/index.jsp Accessed on August 23, 2006
[34] http://www.allexperts.com/ Accessed on August 23, 2006
[35] http://answers.google.com/answers/ Accessed on August 23, 2006

online searching. "Ask a question. Set your price. Get your answer."			access to experts.		Google. Limited domain expertise may be offset by search skills. Experts rated by questioners
Google Co-op[36] "is a platform which enables you to use your expertise to help other users find information."	An expert directory is not directly available.	Topics defined by contributors	Browse or Search for topics to identify topic owner and key contributors	Limited textual description	Self-declaration. Internal Review; community ratings implicit e.g., via # of subscribers
Mad Scientist Network[37] "…an interactive science teaching and community outreach tool"	None	General FAQ provides limited groupings	Browse topic categories, search archived QA pairs as basis for locating experts	Topic areas from fixed taxonomy, textual description, and affiliations,	Self-declaration. Adjudication unknown
Expertise Search[38] "It is our mission to make it easier for consultants and companies throughout the world in finding each other."	Experts organized by domain	18 topic domains	Browse experts lists within topics; search for experts by topic label, region, using keywords	Self-declared profile: Name, Area of Expertise, Specialization, experience, languages, and availability.	Self-declaration. Adjudication unknown

Table 2-3: Representative CoIs supporting Expert Finding

2.4.2 Enterprise Systems

Expert finders are becoming increasingly important in large, heterogeneous organizations. They provide users with capabilities for finding expertise and related information such as published documents, messages, and other information artifacts. Enterprise expert finders may be modeled as end-user applications built upon existing information and knowledge management services. For example, the *Expert Locator* prototype was implemented as a specialized search application built on top of workflow, communityware, and information retrieval services.   The coupling between expert finder functionality and underlying enterprise services provides for multiple assessment perspectives; ranging from operational cost to client-side search support.

[36] http://www.google.com/coop Accessed on August 23, 2006

[37] http://www.madsci.org/ Accessed on August 23, 2006

[38] http://www.expertisesearch.com/default.htm Accessed on August 23, 2006

2.4.2.1   Design Space

Enterprise systems may be described from a number of perspectives including functionality, user interface support, system interoperability, and scalability. These factors and others such as licensing costs, vendor stability, and market share are often critical to whether the system will be procured, successfully integrated into the host environment, and useful to users. While most organizations take an enterprise life cycle view as the basis for introducing expert finder services, the focus here is narrowed to include selected design and performance characteristics useful in discriminating amongst the various commercial systems. A number of systems, representative of the current marketplace, are discussed below. Each system is overviewed in the following areas[39]:

- Philosophy: addresses system organization as to whether explicit or implicit expertise representation schemes are used. Two representation schemes are considered:
    - Pre-coordination schemes involve creating expertise profiles as a precursor to retrieval. Profiles may be generated through self-assessment or through automated analysis of artifacts and social evidence.
    - Post-coordination schemes involve on-the-fly analysis of evidence typically performed as a post-retrieval operation. The two approaches may differ significantly in terms of overall system architecture, retrieval throughput and effectiveness.
- Evidence Sources: work components used to extract evidence. Two classes are considered:
    - Artifacts: viewed as residue of work activity and may be attributed to a particular actor. Artifacts may consist of text (e.g., documents, email, and briefings), images, audio, video, and other object types.

---

[39] The following descriptions draw heavily from publicly released vendor product information or sources internal to MITRE; no proprietary or company-sensitive information has been used.

- Social/Organizational: behavioral information associated with a particular work context or activity. This may include information access patterns, project interactions, conference attendance, and other work behaviors used to identify links between experts. .

- Expertise Model: supports evidence combination and source weighting as a basis for expertise scoring.

- Access Methods: search or browse capabilities used to support expert finding.

- Results Output: the types of output forms used to include:
  - Ranking: expert ordination based on a particular scoring method.
  - Visualization: options to view experts spatially with respect to organizational structure, topic links, or associated activities.
  - Supporting Evidence: evidence used by the system to score expertise scoring.

2.4.2.2 SAP[40]

SAP Expert Finder is integrated into mySAP Human Resources services. This system allows any employee to search for experts stored in user profiles or various text sources such as job postings or job qualifications. Expertise profiles consisting of skill descriptions, experience areas, and task assignments are generated through self-assessment and reviewed by supervisors prior to publication. To bootstrap profile generation, employees may be assigned to one or more work communities which provides a community-specific "template" for entering skills descriptions and selected personal data. Stored profiles can be searched using keyword, Boolean, and proximity searches; however, user search scenarios tied to various community types, may be used to constrain the search interface. For example, an administrator may only be able to search on "name" and search results may be tailored to the user type so that, in this case,

[40] www.sap.com Accessed on August 30, 2006.

the administrator may view an employee's telephone number, fax number, and e-mail address, but no data on the person's expertise areas.

SAP Expert Finder displays a hit list of identified experts. Users can go directly from the hit list to a detailed display, for example, Figure 2-9. The output display and hit list can be tailored to reflect a specific community to which the employee belongs. Table 2-4 provides a summary of the core capabilties of SAP Expert Finder. As indicated it is primarily a database application centered on expert's self-assessment with adjudication by supervisors.
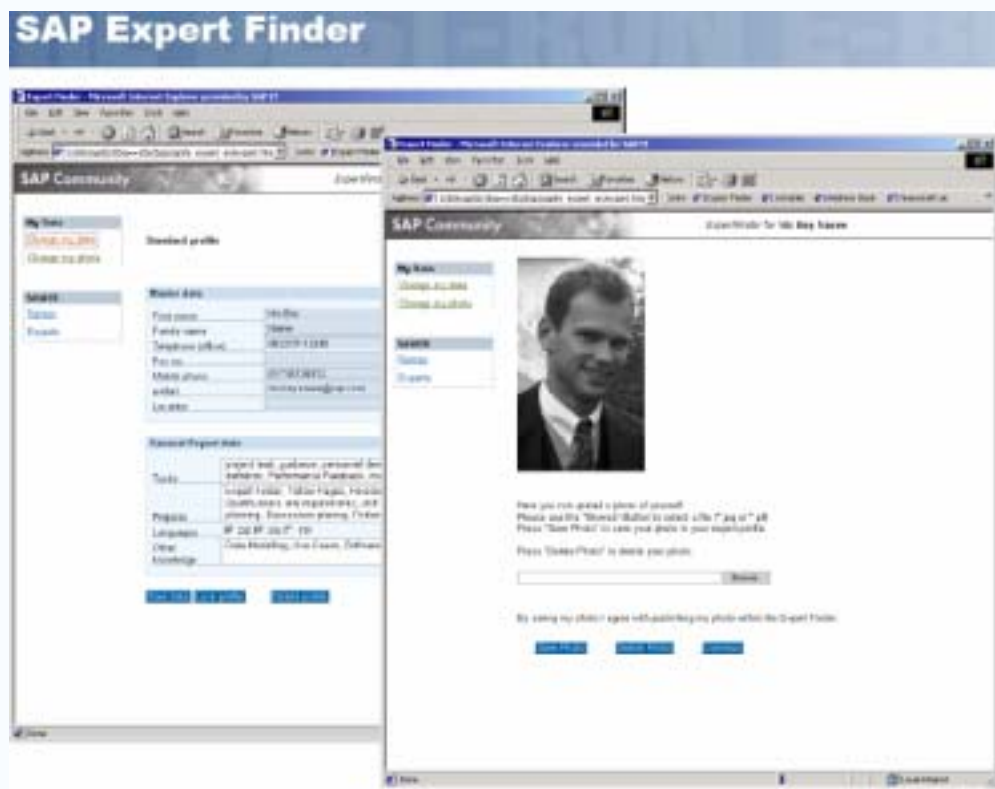
Figure 2-9: SAP: Expert Finder

System Characteristic	SAP Expert Finder Description
Philosophy	**Pre-coordinated**: Expert (i.e., agent) profiles are generated by users as part of a self-assessment process tailored to specific communities the user may be resident within. **Post-coordinated**: Various text objects may be indexed and used to augment profile searches.
Evidence Sources	**Artifacts**: Processes selected text objects; e.g., job postings. **Behavioral:** None
Expertise Model	A formal expertise model is not supported; database matches or relevant text items retrieved are used to identify experts.
Access Methods	**Search**: Standard text and Boolean queries **Browse**: None
Results Output	**Ranking**: None. **Visualization**: None. **Supporting Evidence**: Retrieved experts are described by the matched profiles and associated text items.

Table 2-4: SAP Expert Finder Characteristics

2.4.2.3 Endeca[41]

Endeca enables expert finding through information retrieval services applied to structured and unstructured data from multiple sources. Various indexing strategies may be used to include use of named entity extractors (e.g., InXight[42], Aerotext[43]) and support for taxonomies. Endeca provides a directory search capability that can be used to both manage skills information assigned through self-registration, and to support expert finder searches. Structured directory search is augmented by a text retrieval capability that provides access to query relevant artifacts which can be used to characterize expertise. Document text as a type of expertise evidence can be augmented with searches against past queries; however, historical search patterns viewed as another artifact evidence type are not used as the basis for behavioral modeling. While named entity extraction is used to support document indexing and retrieval it is not used to support

---

[41] http://endeca.com Accessed on August 24, 2006

[42] www.inxight.com Accessed on August 31, 2006

[43] http://www.lockheedmartin.com/wms/findPage.do?dsp=fnec&ti=100 Accessed on August 31, 2006

author identification; essentially Endeca uses available metadata to associate documents to authors.

Endeca does support "Guided Navigation" which provides various filters to reduce a standard retrieval set to meet additional user criteria; essentially working as post-retrieval refinement. For example, a first level retrieval for the query "data mining" may be reduced further by filtering on candidate expert's geographic location, or other attributes through use of dynamic menus which are specific to a particular topic domain or business area.

Relevance ranking based on document-query similarity provides a basic ordering of retrieved items. However, users can filter or sort the list based on other criteria to include business priorities such as geographic area associated with the candidate expert, salary, and other "fixed" characteristics of the expert. The list could also be ordered based on work context to include expert's availability, and project experience. Finally, Endeca's presentation API supports the use of high-level business rules that provide a basis for applying a complex set of conditions to a retrieval list. For example, a user could use a rule to "Highlight the 3 lowest-cost consultants in India that match any criteria the user searches/filters by". Figure 2-10 illustrates search results for the query "Zinfandel" along with "guided navigation" filtering using metadata like price, location, and year. A summary of Endeca's key characteristics is provided in Table 2-5.
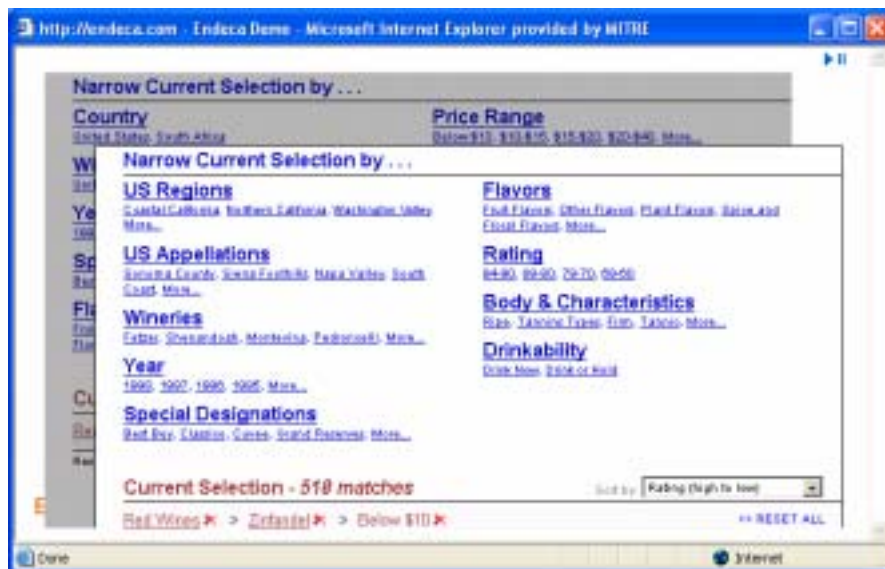
Figure 2-10: Retrieval Results with Guided Navigation used to "drill down".

System Characteristic	Endeca Description
Philosophy	**Pre-coordinated**: Limited skills directory supported. Users self-assess skills in typical applications. Post-coordinated: Text search provides a basis for evidence extraction and relevance ranking.
Evidence Sources	**Artifacts**: Processes a wide range of document types. **Behavioral:** None
Expertise Model	Primarily uses document similarity matching as a surrogate for expertise rating.
Access Methods	**Search**: Standard text and Boolean queries **Browse**: Guided Navigation provides a flexible post-retrieval filtering capability used to "explore" various search subsets.
Results Output	**Ranking**: statistical ranking using query-document ranking model. **Visualization**: None **Supporting Evidence**: Text items are presented along with ranked experts.

Table 2-5:   Endeca Characteristics

2.4.2.4   Tacit ActiveNet[44]

Tacit ActiveNet is similar to Endeca and several other systems that use text retrieval to identify candidate experts. Essentially, published documents and email message text are indexed automatically as a basis for identifying various topics associated with individual actors. For a given query, the system can be used to identify relevant items and associate them with a candidate actor based on authorship. ActiveNet does not exploit email header information (i.e., sender, recipients) and therefore does not exploit social network information associated with email graphs as a basis for identifying experts and expert groups. The system has a privacy model that allows user to build both public and private profiles representing their skill areas and control which information is made visible to users performing expert finder searches. For example, a user can search against both private and public profiles but is not given accessed to profile owners for matches against private profiles. ActiveNet provides protocols for brokering potential contacts between the searcher and retrieved experts that protects expert's privacy where

[44] www.tacit.com Accessed on August 27, 2006

private profiles provide the basis for a match. As with most other commercial systems, authorship is gleaned from available metadata; automatic author identification is not performed.

ActiveNet provides standard keyword and Boolean search support; however, phrase searches are performed against noun phrases extracted by the indexing subsystem. More general phrase analysis is not supported. Retrieval output is presented as a ranked list of experts ordered by confidence ratings based on term frequency and item currency; essentially new documents are viewed as more "valuable" than older items and the system uses an "aging" function to decay document value.   A representative ActiveNet retrieval result is presented in Figure 2-11 and a summary of ActiveNet characteristics is found in Table 2-6.
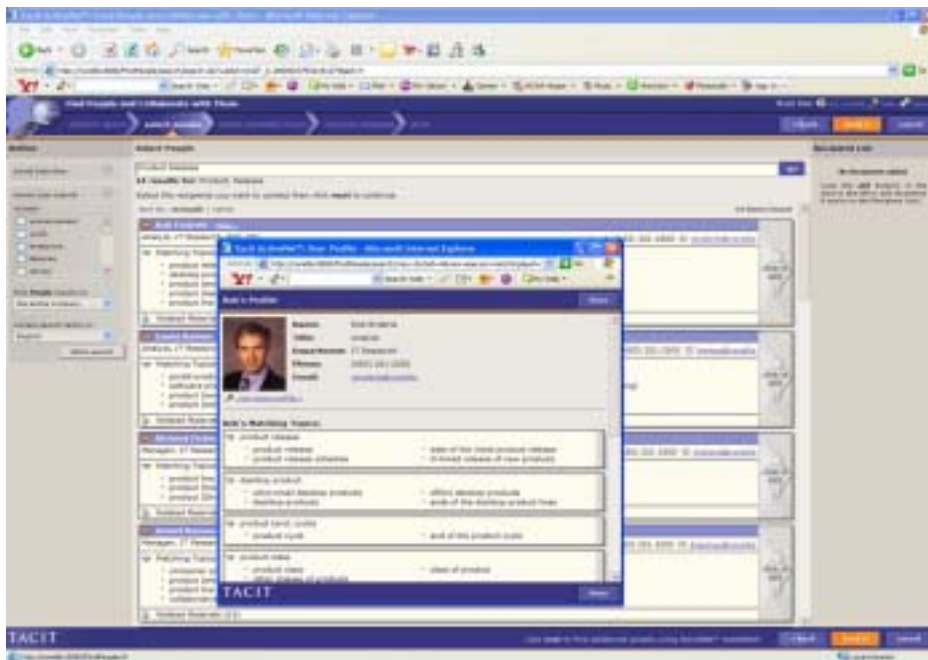
Figure 2-11: Tactit ActiveNet™

System Characteristic	Tacit ActiveNet Description
Philosophy	**Pre-coordinated**: None Post-coordinated: Retrieval operations provide the basis extracting evidence of expertise.
Evidence Sources	**Artifacts**: Processes a wide range of document types to include email. **Behavioral:** None
Expertise Model	Primarily uses document similarity matching as a surrogate for expertise rating.
Access Methods	**Search**: Standard text and Boolean queries **Browse**: None
Results Output	**Ranking**: statistical ranking using query-document ranking model. **Visualization**: None **Supporting Evidence**: Supporting text items are presented along with experts.

Table 2-6: Tacit ActiveNet Characteristics

2.4.2.5   *TriviumSoft[45]*

Triviumsofts's SEE-K is a skills management tool that uses cluster analysis to automatically identify skills areas and associated actors. The Estimation Module provides standard keyword and phrase extraction without the need for lexicons, dictionaries, or skills categories to be defined in advance.   According to available product literature, the phrase extraction methods were general and not tailored to any particular area of expertise. While this makes the system somewhat robust to variation in topic domains it does suggest the system or user must manage indexing "noise" where phrases may not be effective for discriminating amongst expertise areas for a particular query. Noise reduction was not discussed in their online literature; however, several screen-captures depicting various systems modes suggest users have the burden to select terms from a list generated from indexed sources or from a particular skills cluster. This manual "filtering" operation places potentially significant burden on users to eliminate non-skill related terms.

---

[45] http://www.triviumsoft.com Accessed on 15 July 2007.

The system does support access to enterprise resource management systems as a basis for enhancing skills descriptions. For example, employee profiles can be built up using information on training courses taken, formal skills descriptions, and project labor tracking. There is an emerging email processing capability; while not yet commercially available it will provide the basis for identifying skills information from email text and links to "experts" based on email header processing. There is no support for social network analysis other than that provided by the skills clusters generated and the system currently does not analyze worker behavior as a basis for identifying areas of expertise.

While noise in skills descriptions may be an issue, skills groupings presented as a "capability tree" provides an interesting view of enterprise expertise. As shown in Figure 2-12, the capability tree, produced from a proprietary mapping algorithm, is used to characterize enterprise expertise areas in three companies (A, B, C). Company A has strong common skill base (the thick trunk) as well as several skill specialties. Companies B and C have progressively weaker common skill areas and increasing skill diversity. Company C, for example, has no common skill areas and, according to TriviumSoft, this company may have problems with building synergy and overall work coordination.
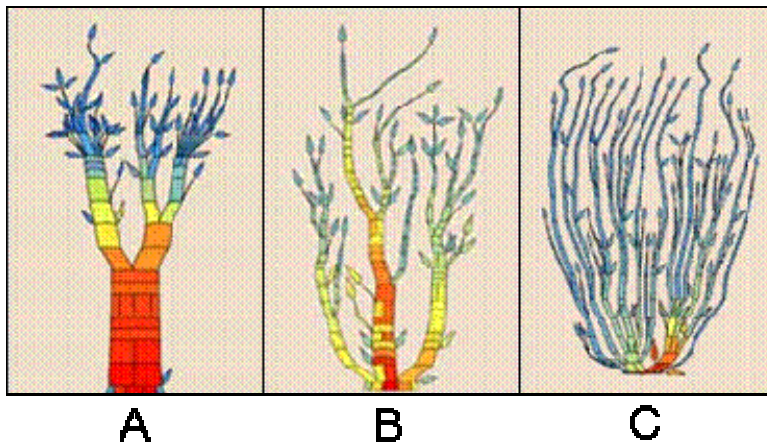
Figure 2-12: Capability Trees for Three Companies

Users can perform full text or Boolean searches. However, the tree structure used to organize skills provides a basis for refining the retrieval list through the addition of new query terms or by

browsing and selection.   A Capability Tree shown in Figure 2-13, depicts a cluster of 600 experts related to "Microsoft/Web Technologies". Color is used to reflect word frequency (red indicates high frequency, blue low frequency). Word importance is reflected by position; center terms are more important than terms on the periphery.
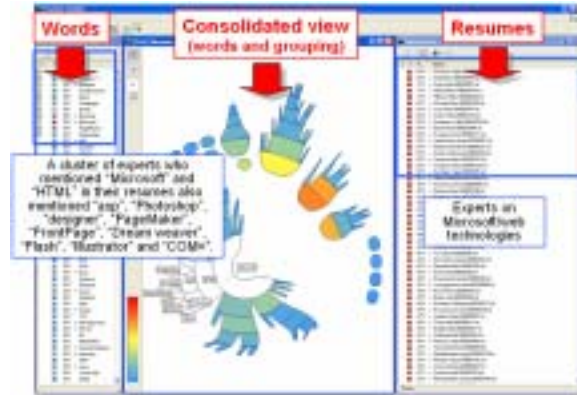
Figure 2-13: TriviumSoft Tree Map Showing a Skills Cluster and List of Experts

System Characteristic	Trivium Description
Philosophy	**Pre-coordinated**: Expert (i.e., agent) profiles may be submitted by users based on self-assessment. Post-coordinated: Expertise profiles and general skill areas are generated through automated cluster analysis
Evidence Sources	**Artifacts**: Processes a wide range of document types. **Behavioral:** None
Expertise Model	Expertise areas are modeled using term significant and co-occurrence. Cluster analysis used to generate expertise skill area models.
Access Methods	**Search**: Standard text and Boolean queries **Browse**: Integrated with search. Users can brows Capability Trees.
Results Output	**Ranking**: statistical ranking based on clustering model. **Visualization**: Capability Trees. **Supporting Evidence**: Text items are presented along with experts.

Table 2-7:   TriviumSoft Characteristics

2.4.2.6 Recommind[46]

Recommind provides enterprise search and categorization tools for a wide range of application domains. For example, Recommind's MindServer is the core retrieval engine supporting the National Library of Medicine's MEDLINEplus[47] site which provides online users access to a wide range of health information. MindServer uses advanced text analysis tools (e.g., probabilistic latent semantic analysis), and categorization tools to identify communities of users based on interest patterns which provide a basis for detecting experts[48]. Essentially, it automatically identifies expertise based on similarity in work artifacts however it can also support user self-declared expertise profiles.

Results are relevance ranked although users can show supporting evidence such as query-relevant documents, person metadata, and project information. A typical expert ranking based on stored profiles is shown in Figure 2-14. Table 2-8 provides a synopsis of Recommind's main characteristics.
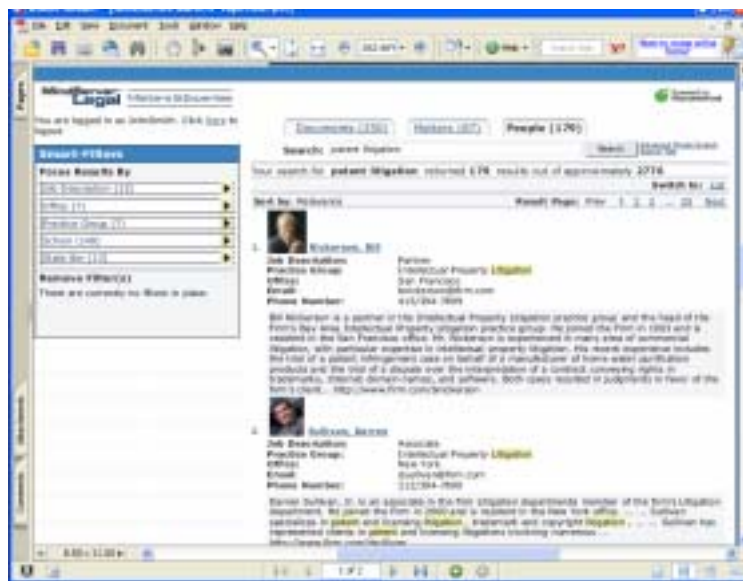
Figure 2-14: Recommind MindServer

---

[46] http:// www.recommind.com Accessed on August 27, 2006

[47] http://medlineplus.gov/ Accessed on August 27, 2006

[48] http://www.recommind.com/index2.php?cat=technology Accessed on August 27, 2006

System Characteristic	Recommind Description
Philosophy	**Pre-coordinated**: Expert (i.e., agent) profiles are by users through self-assessment operations and treated as documents. Post-coordinated: None
Evidence Sources	**Artifacts**: Processes a wide range of document types. **Behavioral:** None.
Expertise Model	Primarily uses document similarity matching as a surrogate for expertise rating.
Access Methods	**Search**: Statistical queries **Browse**: None
Results Output	**Ranking**: statistical ranking using query-document ranking model (PLSA). **Visualization**: None **Supporting Evidence**: Documents, people descriptions, projects and related activities relevant to the query.

Table 2-8: Recommind Characteristics

2.4.2.7   Autonomy[49]

Autonomy IDOL K2 is a full-fledged enterprise search system that processes a range of text (e.g., publication documents, Web pages, briefings, resumes) and email. Verity and Autonomy merged in December 2005 and their legacy products K2 and IDOL were combined into the new product IDOL K2.   From an expert finding perspective, Autonomy has a number of enabling technologies that could be used to support expert finding operations to include advanced statistical pattern matching and Boolean search capabilities, cluster analysis of retrieved items, and a number of tools to add "semantics" to text. For example, entity extraction tools can be used to extract person names, geographic locations, and other elements of text automatically, while taxonomies can be used to provide domain-specific context to indexed text.   While these tools may be used to support expert finding using sophisticated text analysis operations that is not the focus of the current product[50]. Currently, the system generates user agents which model users' information interests and user system operations (e.g., document search). Agents which are effectively dynamic profiles are matched to content as a basis for expert finding. The IDOL

---

[49] http://www.autonomy.com/content/home/index.en.html Accessed on August 24, 2006

[50] For example, Autonomy does not automatically determine document authorship; for example, using entity extraction and other context. Instead it relies on available metadata attached to items. .

Server accepts content and returns similar agents ranked by concept similarity. With that, Autonomy returns a list of individual experts which can be used to obtain contact information. A summary of expert finding characteristics is provided in Table 2-9, below.

System Characteristic	Autonomy Description
Philosophy	**Pre-coordinated**: Expert (i.e., agent) profiles are generated automatically prior to retrieval based on user behavior (e.g., searching) and related content. Agent profiles are separate indexing objects. Post-coordinated: None
Evidence Sources	**Artifacts**: Processes a wide range of document types. **Behavioral:** Over 300 user operations are captured and used to profile user interests; operation context is not considered.
Expertise Model	Primarily uses document similarity matching as a surrogate for expertise rating.
Access Methods	**Search**: Standard text and Boolean queries **Browse**: None
Results Output	**Ranking**: statistical ranking using query-document ranking model. **Visualization**: None **Supporting Evidence**: Text items are presented along with experts.

Table 2-9: Autonomy Characteristics

2.5   Expertise Models and Enabling Technology[51]

Early expert finders were built around core database services in which expertise was captured through a registration process. Systems such as HelpNet, Maron et al (1986), parallel systems like the Dataware II Knowledge Directory[52], in which experts self-nominate, or create skills profiles, stored in a searchable directory. Expertise is accessed through a database query or by browsing experts listed under a specific category heading.   However, expertise profiles can be problematic when built from self-assessment, Shrauger and Osberg, (1981). For example, Davis et al (2006) found that in physician's self-assessments, there was little correlation between self-ratings and actual competency as measured by external assessment. Several studies found that the worst performance was by the least skilled physicians or those with the most confidence.

---

[51] This section draws, in part, from this author's publications on expert finding and knowledge management.

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

Larres et al (2003) found that when measuring computer literacy, students significantly overestimated their actual computer skills. These studies are representative of a wide range of assessments from multiple domains; however there is confounding as to contributing factors. There are cognitive factors that may contribute to self-assessment biases. For example, from the perspective of expertise profile accuracy or completeness there is the potential to not only over state, but to under represent. Some errors in self-reporting may be intentional and relate to cultural or "demand characteristics", Allen and Velden (2005). Here, the self-assessor may be responding to characteristics of the assessment or environment; for example, responses may deviate from "true" answers to hide skill deficiencies or to mask actual capability due to need-to-know or privacy concerns. Altogether, this suggests that in practice registries are not easily scaled especially in large, heterogeneous environments. From an organizational perspective registries may be difficult to populate and update, affecting overall topic coverage[53]. Difficulties aside, expertise registries are still used widely in knowledge management environments where they are integrated with document retrieval, workflow, and other support functions; as discussed earlier in this chapter. However, there is an increasing trend towards automated expert finders that while able to leverage self-assessments and peer ratings, base expertise ratings on behavioral evidence collected automatically from varying work settings.

The shift to automated expertise detection and tracking systems has given rise to a new class of search engines known as *expert finders,* Yimam, (1999). Expert finders are architecturally similar to standard retrieval systems, and align with formal retrieval models; for example, {*D, Q, F, R*}, Baeza-Yates and Ribeiro-Neto (1999); where *D* is a document representation; *Q*, a query; *F*, a framework that associates queries with documents; and *R*, is a ranking function which assigns a score to the similarity between a query and a document representation. However, here, the IR model is recast as an expertise search model {*E, Q, M, R*} where *E* is a source of evidence, *Q*, a query, *M*, a framework for aggregating expertise evidence, and *R*, a ranking function. While IR and Expert Finder models are inherently similar, there may be fundamental differences in their instantiation; for example, expertise evidence, *E*, is viewed as a

---

Accessed on June 18, 2003.

[53] MITRE internal efforts to develop a "skills database" parallel industry experience. The effort was abandoned due to the difficulty of maintaining expertise profiles and, at that time, concerns about privacy.

generalization of documents, *D*, allowing for a potentially wider range of evidence types. As such, in this thesis, IR models are taken as specializations of expertise models. It is noted that this IR perspective can be divided further into retrospective and prospective search models. For example, Balog, Azzopardi, and de Rijke (2006) address use of search models and user profiling methods; where profiling is used to characterize an expert. The notion of generating expertise profiles as first class objects is not addressed directly in this thesis; however, social profiles instantiated as a type of personal network graph are generated as the byproduct of an *Expert Locator* search; this is covered in Chapter 8.

In practice, expert finders are often built around traditional IR systems so that the development of IR-based expert finders parallels the evolution of text retrieval systems. This is evidenced, in part, by recent TREC developments; in particular, the Expert Search task within the Enterprise Search Track (TRECENT) [54]. Here, expert search is focused on finding experts associated with a given topic. The TRECENT 2005 experiment, for example, involves 331,037 documents and 1092 candidate experts selected from the W3C[55]. In this the first effort, search topics were derived from W3C working groups and experts were, by default, group members. This clearly sidestepped potentially complex relevance issues and suggests tempering current results somewhat at least until a more extensive relevance-assessed topic set can be developed. While the W3C collection provides a useful basis for evaluating certain aspects of expert retrieval, it is somewhat limited; at least when compared with the diversity of sources and work context associated with operational environments. In TREC 2006 a new 55 topic test set with relevance judgments provided by TREC participants has been developed[56].

TRECENT addressed a number of issues regarding collection, indexing, retrieval, and expertise ratings. Figure 2-15, views this from an architecture perspective, reflecting functional areas associated with a number of reported efforts. Most participants exploited multiple collection

---

[54] The TREC Enterprise Track, http://www.ins.cwi.nl/projects/trec-ent/, "has as primary goal to build a test collection for *Enterprise Search*. Enterprise search considers a user who searches the data of an organisation in order to complete some task. Enterprise search is interesting because it has not been sufficiently addressed in research, and it is of immense practical importance in real organisations."

[55] The World Wide Web Consortium (W3C), http://www.w3c.org Accessed on July 15, 2006.

[56] http://www.ins.cwi.nl/projects/trec-ent/wiki/index.php/Main_Page Accessed on September 11, 2006.

sources. MacDonald et al (2005), viewed documents from multiple sources as separate evidence types and allowed documents to be weighted by type. With that, emails may be given different weight than, say, homepages. Azzopardi et al (2005) focused on email discussion lists; Yao et al (2005) used emails in combination with documents and personal or organizational homepages as entry pages. In selected cases, document structure was exploited directly. In particular, some researchers used document submodels based on type as the basis for refining entity extraction and similarity computations. For example, Cao et al (2005) developed window-based submodels based on document metadata such as <Author>, <Title> and <Body>. These models were used separately or in various combinations to score experts as part of the retrieval operation. For example, query terms co-occurring with topic terms in the <BODY> may be treated differently than if occurring in different "fields".
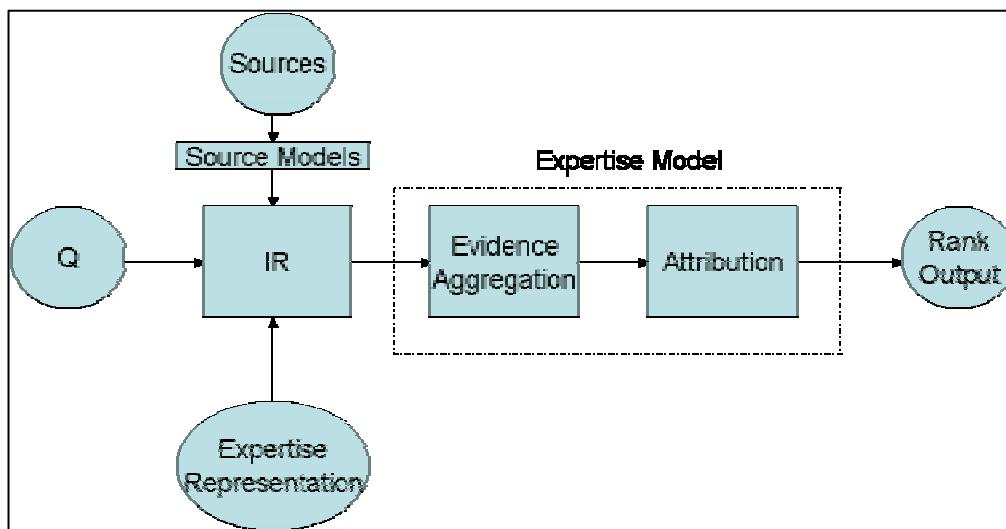
Table 2-10: Conceptual Architecture

Regarding expertise representation, most systems generated expert profiles as a type of composite document. While this representation scheme is clearly document-centric it supports a wide-range of IR models. Zhu et al (2005) used entity extraction to identify person names that formed queries against the document collection. Documents containing each name were used to build a document grouping; treated as an expertise profile. MacDonald et al (2005) used up to three sources (expert's homepage, documents containing expert's name, and emails) to build expert profiles. Fu et al (2005) use document "reorganization" to build composite documents

70

used to characterize an expert's expertise. In particular, various sources are processed to extract information relevant to a particular expert; source-specific rules are used to normalize text and merge into a composite document representation.

While most participants implemented profile generation as a pre-coordination, Yao et al (2005) extracted candidate experts and performed expert ranking as a post-retrieval process. This parallels a number of earlier research systems, for example, Mattox et al (1998, 1999) use entity extraction as a post-retrieval process to identify candidate experts mentioned in retrieved documents. Essentially, for a given query, document relevance provides a basis for conferring expertise on authors or other persons; to include named entities (i.e., persons or groups) embedded within the text. This effort is interesting in that it led to one of the earliest known systems that integrated an autonomous search engine into the overall approach. Here, an enterprise search engine supported the first pass retrieval operation and entity extraction was used to assign "experts" (i.e., authors or mentions) to retrieved items. This deviates from current TREC systems in that the Mattox expert finder, due to policy restrictions, did not have access to the underlying document (i.e., artifact) collection. This precluded a priori profile generation based on access to the corporate search engine index, the raw collection, or on independent "crawls" of the corporate Intranet. D'Amore (2004, 2005) extended this model as part of a distributed search architecture in which multi-evidentiary sources (to include project data and organizational ratings) are accessed consistent with processing costs, and corporate policy constraints.

A range of IR models were used in TRECENT. A common approach was to use language models to assess document relevance to a query; Azzopardi et al (2005). Other approaches factored in the probability that an expert was correctly identified in the target document. This 2-stage model effectively juxtaposed a language model with a co-occurrence (i.e., attribution) model. For example, a language model used to assess the probability that a query was generated from a specific document model, was multiplied by the probability that a document is associated with a particular entity; Azzopardi et al (2005). A number of other systems used the traditional Vector Space and Latent Semantic Indexing methods. However, regardless of the similarity

model used, identity resolution (accurately linking, say, an author to a relevant artifact) is central to overall performance.

More generally, identity resolution (or attribution) is central to a number of problem domains such as social network analysis, terrorist screening, border control, and criminal investigations. It primarily addresses the issue of discerning an actor's identity from multiple instances where identity may be confounded due to name variants, missing attributes, and relationships and activities. For example, an actor's identity "signature" may consist of name components extracted from email sender field and signature block, or "mentions" extracted from various documents. Names may include legal variants, misspellings, and nicknames which may be useful discriminators or otherwise confound one actor with another. Identity elements may also include attributes such as titles ("Chief Scientist"), office location, and other features used to discriminate one actor from another. The identity resolution problem is ubiquitous; for example, Esayed and Oard (2006) have addressed the problem in email archival. It is also a key problem in expert finding as reinforced in several TREC papers. For example, several TREC systems either exploited email signature blocks or header information to extract name elements or used "mentions" extracted from various documents. Ru et al (2005) used simple rules to handle homonyms; Zhu et al extracted names from emails headers, and Azzopardi et al used various match levels to identity name variants. Match types included: exact match, and match on Last Name and Initial. Ru took a similar approach but used several heuristics to filter out name elements that were deemed to be ineffective to include "short" components (less than 3 characters) and common names like "Tom". In practice entity resolution is a critical problem and methods such as named entity recognition may be insufficiently accurate to work in high-precision environments without extensive post processing.

Systems, centered on documents as evidence of expertise may not fully exploit additional work context and relationships. As noted, MacDonald et al used local document weighting based on source as a proxy for work context weighting; presumably this scheme could be extended to apply to specific document type in multiple work settings. That is, an email in one discussion list may be weighted separately from an email in another discussion list. In addition, Cao et al used cluster ranking to enhance search results in that relations between people were used to

modify ranks so that people who appear in similar (semantic) contexts or who co-occur in the same artifact may receive adjusted weights. These methods are representative of a wider range of methods emerging that exploit work context more directly; especially where coupled with extant workflow and productivity tools found increasingly in many enterprises. For example, Productivity enhancement environments such as Sharepoint[57], Lotus Notes[58], and IBM Workplace[59] are providing the infrastructure needed to improve work visibility and expertise awareness. These environments provide necessary tools and infrastructure needed to manage diverse communities-of-practice centered on technology areas, projects, and various other business activities. For example, Microsoft's Sharepoint provides services for contacting community members using e-mail and instant messaging; content management, and site personalization used to tailor user views and information access to include automatic alerting mechanisms. Here expertise detection may simply involve identifying a community related to a particular expertise area, identifying community members, and confirming relevance through analysis of member interactions and work artifacts such as briefings and whitepapers. In particular, Sharepoint was one of several *activity spaces* exploited by the *Expert Locator* developed in this thesis.

Activity-centric work environments provide a potentially rich context for *mining* expertise. The notion here is that artifacts, social interaction, and activities relevant to a particular domain may serve as expertise indicators when viewed from the perspective of a particular expertise model. For example, Autonomy Agentware Knowledge Server[60] analyzes users' search and publication histories to determine concepts that are indicative of their expertise. Yenta (Foner, 1997) determines user expertise from email message traffic, as does Tacit KnowledgeMail[61]. KnowledgeMail does not exploit email routing information to identify experts and this is consistent with their privacy model. Alternatively, Schwartz and Wood (1993), describe a system that uses the directed graph obtained from e-mail message headers to find affinity groups

[57] http://office.microsoft.com/en-us/sharepointtechnology/ Accessed on July 22, 2007.

[58] http://www-306.ibm.com/software/lotus Accessed on October 16, 2006.

[59] http://www-142.ibm.com/software/workplace/products/product5.nsf/wdocs/workplaceoverview Accessed on October 16, 2006

[60] http://www.autonomy.com/tech/wp.html Agentware Knowledge Server. Accessed on October 15, 2006

[61] http://www.tacit.com/products/knowledgemail.html Accessed on July 15, 2006

without using message text. Wang et al (2002) developed an expert finding algorithm that used user browsing captured through web log analysis to infer expertise. The more high quality web pages a user has visited the higher the assessed expertise level.   The current approach does not use page content to assess quality; it uses a modification of the HITS algorithm Kleinberg (1999) to determine page and user importance scores.

The Bellcore Advisor (also known as *Who Knows*), was used to find people with explicit expertise in a 5000 person company, Streeter and Lochbaum, (1988). Here various research groups were characterized by descriptions of projects and other activities. These groups were represented using automatically extracted terms and Latent Semantic Analysis (LSA) was used to represent both groups and terms in factor space. While the focus was on retrieving people with certain expertise, the system retrieved research groups that were "close" to a query. McLean et al (2003) developed PeopleFinder, a Web-based system which automatically identifies experts based on published documents. The prototype leverages organizational data such as project descriptions and membership to infer which documents can be used as evidence of expertise. This system supersedes earlier work on P@NOPTIC Expert, Craswell et al (2001), which had limited performance in part due to the use of "low-quality" documents as evidence. PeopleFinder addressed that issue by linking candidates to other, highly relevant and unattributed documents based on *proximity*. For example, a project member could inherit unattributed relevant documents collocated within the same project space. A similar approach was used by D'Amore et al (2003) where projects were modeled as task hierarchies, so that task members could be assigned to documents they were explicitly associated with as well as to unattributed documents based on various task "closeness" measures. In addition, both Craswell and D'Amore used document type as a basis for weighting importance; for example, a project page or a home page may be more relevant than say a news page. As noted, above, several TREC systems adjusted evidence weights according to document type.

Expert finders may be integral to specific application domains. For example, Becks et al (2003) discuss an expert finder capability integrated into an e-learning environment. This system is designed to make "co-learners aware of each other". In contrast to expert finders that exploit artifacts as the main sources of expertise, they use a dual approach that exploits evidence of user

experience, such as project work, along with the interaction history between user and course material. Becerra-Fernandez (2000) describes Expert Seeker a people finder application used to locate experts at the National Aeronautics and Space Administration (NASA). The system exploits corporate-wide data as the basis for augmenting user-specified profiles. Self-assessment provides a starting basis for generating expertise profiles; from this various other sources are used to augment descriptions in a type of bootstrapping approach. For example, data from Human Resources databases, director services, and skills databases are used to enhance manual skill entry and identify user contact networks. In addition, they use an employee performance evaluation system to qualify skill level; without discussion of privacy implications.   Other relevant data such as hobbies, project membership, civic activities, and employee picture are also used in the employee expert profile. As with other systems, they augment self-assessments and corporate data with document analysis to identify additional context for areas of expertise.

Few systems have been designed to more directly exploit social network information as a basis for expert finding. For example, CORDER[62] is a relation discovery capability that identifies affinity groups based on common characteristics such as shared topic interest (i.e., expertise), co-work, and contacts, Zhu et al (2005). For example it may be useful in generating a buddy list for instant messaging use based on overlapping publications. It uses named entity extraction to identify individuals within documents and other artifacts as the basis for computing association matrices which define relationships between actor-actor pairs. Each matrix may describe a relationship between any pair of actors. The system uses representative topics to characterize organization interests or expertise areas so it may not necessarily extend to support ad hoc expertise queries. Bao et al (2006), exploit multiple relationships within a single framework. In particular, given two sets of objects, for example experts and expertise areas, the Typed Separable Mixture Model (TSMM) uses all types of co-occurrence information with a single model. Here expertise may be assessed based on the relationships between experts, and between topics as well as the ties between experts and topics; overall these relationships may be instantiated as single mode and bipartite graphs.

---

[62] CORDER: COmmunity Relation Discovery by named Entity Recognition

The focus on social network structure suggests viewing expert finding as a type of social matching problem, e.g., "systems bringing people together", Terveen and McDonald, 2005. While this includes expert finders that leverage email routing lists, project membership, and organizational structure it also admits community-based services and *Personal Information Management* (PIM) systems whose primary focus is not ostensibly expert finding. For example, ContactMap, Nardi et al (2002), and Whittaker et al (2004), is built on a social desktop metaphor allowing users to organize contacts according to user-specified relationships. Broadly viewed it supports a type of shared workspace providing various "social" cues used to coordinate work and, potentially, to access needed expertise. While ContactMap is not specifically oriented to the problem of expertise detection and tracking, its overall design is suggestive of how users may generate expertise maps as a type of personal network in which users can exploit contact history and expertise characterizations as the basis for identifying specific experts. Here, users may follow personal network links to identify contacts that may provide referrals, or that may satisfy an expertise need directly. More broadly, expert finding is viewed increasingly as a derivative of social interaction and expertise exchange.

Social bookmarking sites such as del.icio.us[63], digg[64], Technorati[65], and StumbleUpon[66] suggest a community framework for expertise exchange enabled by user-controlled resource tagging. Social tagging extends traditional browser-based bookmarking such as supported by Internet Explorer and Firefox by storing bookmarks in a centralized store easily accessible from different access points. Users can share bookmarks with others and browse or search them based on user-defined topic tags or free-text annotation. This social tagging provides the foundation for simple expertise exchange such as provided by *Cogenz*[67]; a system that supports "identity" tags through a self-declaration process. Users can build personal profiles and assign tags which can be used to support a simple expertise search or to potentially link profiles. The MITRE-developed *Onomi*[68] system Damianos et al (2006) provides enterprise social bookmarking designed to

---

[63] http://del.icio.us/ Accessed on 17 July, 2007.

[64] www.digg.com Accessed on 18 July, 2007

[65] www.technorati.com Accessed on 18 July, 2007

[66] www.stumbleupon.com Accessed on 19 July, 2007

[67] http://www.cogenz.com/ Accessed on 17 July, 2007.

[68] www.mitre.org/work/tech_papers/tech_papers_06/06_0352/06_0352.pdf   Accessed on 18 July 2007

facilitate sharing across research areas, social network formation around topical interests, which will "feed expertise finding and user profiling". Here, topic tagging provides a basis for identifying experts based on tags usage and for expertise profile generation in areas not aligned with standardized corporate taxonomies. *Dogear*, Sastry (2006), addresses the problem of expertise exchange across security firewalls or other organizational boundaries. A hashing scheme is used to capture the relationship between expertise descriptors (keywords, typically) and a user identifier; There is the potential for "noise" matches[69] and limitations on the complexity of expertise indicators provided; however, the main focus of the system is less on the underlying expertise model and more on expertise transfer across distributed environments. The system is built around an "expertise dictionary" (ED) which is a representation scheme used to attribute network IDs (essentially, users) to expertise indicators. The system trades off cross-organizational interoperability for potentially reduced retrieval performance. For example, *Expert Locator* supports client-side browsing using interactive visualization; here, an ED would likely be insufficient for rendering the rich navigation space necessary to facilitate iterative search and end-user browsing across a corporate Intranet. However, the notion of providing expertise exchange mechanisms in support of cross-boundary expert finding is sparking interest in areas such as the Semantic Web[70].

Cross-boundary social tagging methods may be viewed architecturally as middleware components; juxtaposed between Semantic Web lower-level enabling technologies and various applications such as collaboration and workflow. For example, the ExpertFinder[71] initiative is focused on leveraging the Semantic Web for creating the infrastructure needed to support expert finding across the Web. This includes vocabularies and rule sets needed to annotate personal home pages, conference pages, publication lists, and other sources with expertise descriptors that can be exploited by expert finder tools and services. For example, Aleman-Meza et al (2006) provides "a framework for the reuse and extensions of existing vocabularies in the Semantic Web." Here, expert finding provides the application focus for reusing vocabularies found

---

[69] The system encodes expertise descriptors using a multiple hashing scheme that may incur collisions.

[70] http://www.w3.org/2001/sw/ Accessed on 18 July, 2007. The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.

[71] http://lsdis.cs.uga.edu/~aleman/efw2007/ Accessed on 18 July 2007. Expertfinder Workshop is sponsored by EU NoE Knowledge Web and the Knowledge Nets project, which is part of the InterVal-Berlin Research Centre for

already in FOAF[72], vCard[73], Dublin Core[74] and others.   The goal is to explore vocabulary reuse across diverse communities to facilitate expertise detection. Iofciu et al (2007) explore methods for extending user-generated FOAF files (used to characterize expertise) with automatically generated profiles. Extended FOAF files, called ExpertFOAF, can be built from wide-ranging resources associated with a particular user or organization. This approach however does not resolve long-standing issues associated with manual profile generation or automatic approaches that extract expertise indicators from various work contexts; especially problematic are instances where composite ExpertFOAF profiles are generated from disparate (e.g., multi-organizational) environments.

FindXpRt, Li et al (2006), goes further in terms of adding new facts automatically (i.e. expert indicators) based on deductive reasoning, and in using rules to enhance expert detection, selection, or referral. The system is used to support collaboration and various rules and taxonomies are used to match a user with a candidate collaboration partner. User-specific facts are extended using a taxonomy to provide a broader basis for characterizing expertise needs and experts.   FindXpRT is built on FOAF; however while FOAF supports person related facts, FindXpRT extends FOAF to support rules using a formal rule language such as RuleML. RuleML markup can be used to address, in part, the *selection* problem describing earlier in this chapter. For example, a contact preference rule may be used to restrict contact to candidate experts based on organizational rank. They use a declarative language POSL[75] to generate facts and rules in a human-readable form; which are translated into RuleML syntax. OO jDREW[76] is a rule engine used to process FOAF rules.

the Internet Economy, funded by the German Ministry of Research (BMBF)

[72] http://www.foaf-project.org/ Accessed on July 20, 2007. "The Friend of a Friend (FOAF) project is creating a Web of machine-readable pages describing people, the links between them and the things they create and do."

[73] http://www.imc.org/pdi/vcardoverview.html Accessed on July 20, 2007. vCard is used to support personal data exchange; for example, business cards.

[74] http://dublincore.org/ Accessed on July 20, 2007. "The Dublin Core Metadata Initiative is an open organization engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models."

[75] http://www.ruleml.org/submission/ruleml-shortation.html Accessed on July 20, 2007. "POSL is a positional-slotted language… that integrates Prolog's positional and F-logic's slotted syntaxes for representing knowledge (facts and rules) in the Semantic Web."

[76] http://www.jdrew.org/oojdrew/ Accessed on July 20, 2007. "OO jDREW is a deductive reasoning engine for the RuleML web rule language (including its OO extensions), written in the Java programming language."

Agent-based systems address the problem of locating expertise where evidence is distributed across a large number of actors; for example, Maybury and D'Amore 2001 discuss agent-based searching used to detect large-scale communities and indications of "common social opinion or concern." Yu and Singh (2001) describe a referral system in which agents using local knowledge represented by expertise profiles, to find experts. A query is propagated throughout the social network, and each agent assesses the query with regard to a stored profile or routes it to other agents based on knowledge of their neighborhood. ContactFinder, Krulwich and Burkey (1995, 1996), is an intelligent agent that runs on on-line bulletin boards. The agent reads questions posted on the bulletin board and responds with referrals to other users that are likely to be of assistance. Expert Finder, Vivacqua (1999), uses a personal agent to profile users. The system was able to exploit knowledge of the Java programming language to build expertise profiles. Jie et al (2000) propose a framework for ontology-based agent system for enterprise expert detection. The hybrid architecture supports local neighborhood searching (peer-to-peer) while allowing for the emergence of a central authority which can support more of a top-down search. The framework allows the use of an organizational ontology (essentially the organization chart) to be used to escalate expertise finding to organization members (e.g., supervisors, mentors). The search process can be constrained by security or privacy conditions built into user agent profiles. However, there is no evidence the system was actually built and a number of challenging problems regarding text analysis and social network construction were not addressed in the paper. Their proposed use of organizational context, however, is consistent with the *Expert Locator* model described in this paper.

2.6   Expert Locator

A wide arrange of systems and expertise models have been discussed here. This includes research systems emerging from formal research environments, such as TRECENT, as well as efforts that are situated within rich organizational environments. Paralleling this are an increasing number of commercial and open community systems focused on providing expert finder services as autonomous offerings or as integrated into enterprise workflow or community

services. While much of the focus is on a target enterprise or group, there is increased focus in cross-boundary expert finding using the Semantic Web and other enablers.

The remainder of this thesis focuses on the *Expert Locator* development. *Expert Locator* is enterprise-centric in its current implementation and explores use of multiple sources of evidence as the basis for characterizing expertise and identifying experts. The system supports standard queries but also provides a visual interface that facilitates end-user navigation through various organizational and social networks as the basis for ferreting out needed expertise.

3   Expertise Signaling

There is an extensive literature focused on *experts*, their characteristics and behaviors. Generally, experts are viewed as high-performers having superior knowledge and problem solving skills when compared to novices; however, this runs counter to what is known about expert's performance in various decision contexts where cognitive biases may contribute to poor performance in decision making or predictive tasks. Yet from this disparity emerges a constant: experts signal their expertise. Experts signal their skills and experience to advertise capabilities, build reputation, and establish trust. Signaling behavior is visible and provides a basis for detecting experts, identifying relevant organizational context, and mitigating the problem of explicit expertise encoding.   The remainder of this chapter explores the nature of expertise, and lays the groundwork for the signaling-based expertise model discussed throughout the remainder of this thesis.

3.1   The Nature of Expertise

Merriam-Webster[77] defines an expert as "having special skill or knowledge derived from training or experience".   While this definition is elegant in its simplicity, it belies the true complexity as reflected in the extensive literature on the nature of expertise. In particular, the definition of expertise varies across studies and environments and is typically determined by those at the center of study, Huber (1999). Definitions may be operationalized and rooted in a particular practice; for example, "an expert radiologist is a radiologist who is able to detect subclinical breast lesions and to precisely locate them within the breast", Coibion (1995). Similarly, expertise studies may be reinforced by underlying behavioral models; for example, Jensen (1995) studied the behavior of pilot decision making and found that expertise was associated with four functional areas: aviation experience, risk management, dynamic problem solving, and attentional control.

---

[77] http://www.m-w.com/info/election.htm. Accessed on November 18, 2002

While expertise may be assigned through observed behaviors, it may also be formally ascribed; especially where there are legal constraints or tests as to what constitutes *true* expertise. For example, the Securities and Exchange Commission (SEC) recently revised legal-driven criteria used to define "financial expert". Here, *financial expert*, originally defined under Sarbanes-Oxley[78], was broadened to shift fiscal responsibility to CEO's and other corporate senior executives. The label *financial expert* was changed to *audit committee financial expert* and includes the following explicit criteria:

- An understanding of financial statements and GAAP[79];
- An ability to assess the general application of those principles in connection with the accounting for estimates, accruals, and reserves;
- Experience preparing, auditing, analyzing, or evaluating financial statements that present a breadth and level of complexity of accounting issues that are generally comparable to the breadth and complexity of issues that can reasonably be expected to be raised by the registrant's financial statements, or experience actively supervising one or more persons engaged in such activities;
- An understanding of internal controls and procedures for financial reporting;
- An understanding of audit committee functions.

The definition of *audit committee financial expert* includes qualifications such as:
- Education and experience as a principal financial officer, principal accounting officer, controller, public accountant, or auditor or experience in one or more positions that involve the performance of similar functions;
- Experience actively supervising a principal financial officer, principal accounting officer, controller, public accountant, auditor, or person performing similar functions, or experience overseeing or assessing the performance of companies or public accountants with respect to the preparation, auditing, or evaluation of financial statements.

---

[78] http://www.sec.gov/about/laws.shtml#sox2002 Accessed on May 8, 2006

[79] GAAP is Generally Accepted Accounting Principles

While most definitions of expertise are situated, some studies have focused more on a general calibration of expertise; one not closely tied to any one domain or work context. For example, Dreyfus and Dreyfus (1986) viewed expertise from the perspective of skill acquisition. They studied skill acquisition across a number of domains (e.g., chess) as basis for defining "Five Stages of Skill Acquisition" which provide a domain-independent scaling of expertise. These included: *Novice*, *Advanced Beginner*, *Competent*, *Proficient*, and *Expertise* (*Expert*). There is a progression from *novice*, who uses facts and rules not necessarily grounded in a particular setting or problem context, to *expert* whose actions are based on intuition and wisdom that is situated and context specific. Similarly, Gaines (1988) views expertise acquisition in terms of skill and experience progression through exposure to problems:

> *The formation of expertise is functional in general because it leads to division of labor in the management of knowledge acquisition. The development of an individual expert is a random process brought about by strong positive feedback loops in the social process; for example, that a proto-expert with superior performance is brought more problems and hence has a greater opportunity to learn and improve that performance. A diversity of such positive feedback processes operate in the professions and sciences with little relation between them except their overall effect in promoting the formation of expertise."*

While the problem of defining expertise is addressed from several vantage points, a universal definition of an expert is lacking and the study of expertise is exacerbated further by lack of consensus on a research framework or guiding principles, Huber (1999). This inhibits transferability of results so that "approaches used by researchers, typically in controlled settings, are unlikely to mirror the assessment of expertise by individuals in applied contexts", Shanteau et al (2003).

This leads also to uncertainty as to use of the term *expert*, as there is often no clear basis for assigning it. It also raises issues as to how expert finder systems can be evaluated effectively within a particular environment or how evaluation results can be compared from multiple studies and disparate environments. On a more practical note, within the organizational setting for this thesis[80], individuals are more often characterized as having certain *knowledge, skill*, or *expertise*

---

[80] The MITRE Corporation, www.mitre.org . Accessed on May 2, 2006

as it stops short of conveying some quantifiable level of mastery. This does not preclude use of the term *expert*, but instead reserves its use for cases where there is a relatively strong consensus and visible evidence as to qualifications and skills. This usually takes the form of formal credentials, skills made visible in a particular business or technical context, and external relationships or activities, such as conference participation and membership in prestigious committees. Regardless of when the term *expert* is used, it is essentially a generalization or label applied to those that exhibit specific behaviors Shanteau (1992).

3.2   Expert Behavior: A Cognitive Science View

Cognitive science research suggests that experts can generally be discriminated from non-experts based on a number of individual characteristics and behaviors. In particular, it takes time for experts to identify optimal problem solving strategies tailored to some domain. For example, Ericsson, Krampe, and Tesche-Romer (1993) note that expertise is acquired over time and associated with increased practice; Chase and Simon (1973) note that it takes about 10 years to become a world-class chess player. Rosenbloom and Newell (1986) describe a power law behavior for skill acquisition in which skill is more rapidly acquired in the early learning phases but is much harder to increase within increasing skill level; a law of diminishing returns. This is especially true when viewing expertise from the perspective of organizational knowledge coupled with, say, technical expertise; there, given diverse work performed across disparate operating units, it is often very difficult to acquire *deep smarts,* Leonard and Swap (2005), and to keep expertise current.

While it takes time to become an expert, experts turn out to be very skillful in optimizing problem solving strategies to a particular domain. Experts use problem decomposition and structuring more effectively than non-experts. Simon and Chase (1973) note that experts employ particular strategies to include acquiring problem solving information in chunks, patterns, and more complex knowledge constructs. In fact, chess grandmasters encode somewhere between 10,000 and 100,000 chunks of information, Simon and Gilmartin, (1973). There is also physiological evidence in eye movement studies, DeGroot and Gobet (1966), where chess experts were able to identify key chess board patterns more rapidly than non-experts. Chi,

Hutchinson, and Robin (1989) focused on the definition of knowledge structure within a specific domain, and the relationship between structure and use. In one study, expert children were able to make specific inferences and perform categorical reasoning based on hierarchical knowledge structure and cohesive local knowledge; something that non-expert children were not able to do. Glaser (1986) was able to show that high levels of competence result from the interaction between knowledge structure and processing capabilities.

Bedard and Chi (1993) assessed the influence of domain knowledge on perceptual processes and strategies in problem solving. Across three clinical problems with varying complexity, a number of second year and third year nurses with high and low academic scores were assessed with regard to their ability to generate hypotheses, identify disconfirming information, and to correctly diagnose the case. While academic ability affected decision making accuracy in low complexity tasks, domain knowledge was a stronger determinant of decision accuracy for more complex tasks.

In general, experts have more effective memories, retain more knowledge than novices, and can call it up more efficiently. Ericsson and Polson (1988) focused on the memory skills of a headwaiter. Expertise in memorizing restaurant orders was associated with five skilled memory characteristics: efficient information encoding, retrieval structures built around encoding schemes, the use of long-term memory for effective retrieval after immediate use, rapid encoding, and domain specificity. The expert was also able to handle orders regardless of the order in which the items were presented. Performance dropped for tables of, say, eight diners where sequence variation slowed order taking somewhat, but even in this case recall and accuracy were not affected. Interestingly, the expert that was tested was able to flexibly and more generally apply his memory skills to other tasks when the tasks had similar structure. He had a number of menu-specific skills for example, schemes for encoding salad orders or how well a steak was to be cooked, that transferred to other tasks related to encoding time and flower names. Superior long- and short-term memory skills were evidenced in children with expertise in video game playing, Vandeventer (1997). Expert game players also exhibited many of the same skills identified by Glaser and Chi (1988) to include domain excellence, identifying large patterns in games more so than novices, and speed in problem solving. However, while there is

some cross-study convergence regarding the relationship between domain knowledge, knowledge structuring, memory skills, and processing methods, the <u>performance of experts on certain tasks is more variable</u>.

3.3   Expert Behavior: A Decision Analysis Perspective

Performance-related research is divided as to how well experts perform on a range of tasks. Research in the decision sciences suggests that experts perform poorly across a number of decision analysis tasks. Experts make flawed decisions and employ heuristics that introduce significant biases in the analysis task. Foss, Wright, and Coles (1975) discussed the low validity of expert assessments in judging livestock, even when compared to novices. In Dawes and Corrigan (1974) experts were shown to under perform simple linear models across a range of forecasting problems. While experts were effective in determining the key variables or factors in the prediction problem they often relied on heuristics and that resulted in a number of biases such as anchoring and availability, Kahneman, Slovic, and Tversky (1982). Experts often do not exploit available information, Goldberg (1970), and this has been reported for court judges, Ebbesen and Konecni (1975), and clinical psychologists, Goldberg (1970).

This discerning view of expert performance is juxtaposed with cognitive science research which suggests that experts are competent and have both knowledge and functional skills that are distinct from novices. The difference in findings between decision science and cognitive research suggests that other factors may be involved.

Shanteau (1992) suggests that the different view of experts held by decision and cognitive scientists is explained by differences in task characteristics. Shanteau presents a "theory of expert competence" that suggest that both analyses are correct but incomplete. He lays out five components of competence, (*sufficient domain knowledge, psychological traits, cognitive skills needed to make decisions, use of appropriate decision strategies,* and *tasks characteristics*) and concludes that the difference between the decision science and cognitive science literatures is related to differences in <u>task</u> domains studied.   As examples he sights the relatively high performance (and internal consistency) of weather forecasters as compared to clinical

psychologists, and stock brokers that have performance that is close to random, see also, Stewart, Roebber, and Bosart (1997).

A number of researchers have looked at task characteristics as a basis for assessing performance. Orasanu and Connolly (1993) identified eight factors that were characteristic of what they termed naturalistic decision making (tasks): ill-structured problems; uncertain, dynamic environments; shifting, ill-defined or competing goals; multiple event-feedback loops; time constraints; high stakes; multiple players; <u>organizational norms and goals</u> that must be balanced against the decision makers' personal choice. For example, *organizational norms and goals* can be viewed as elements of *culture*; that is, "a body of learned behaviors common to a given human society.[81]" Culture, as a behavior shaping mechanism, affects human performance in wide-ranging settings. Heuer (1999) links culture to analytical bias associated with one's own self-interest, organizational setting, and social norms. For example, in an intelligence organization, culture may bias information sharing practice and dictate product coordination across expert forums, analysis domains (e.g., INT's),[82] and organizations[83]. In affect, expert performance is affected by multi-organizational cultural biases as to how information is shared across organizational boundaries, integrated, and aligned with policy. For example, Davis (2001) identifies key policy changes that occurred as the result of the Gulf War, the Balkan crises, and 11 September terrorist attacks and their affects on information sharing and coordination. In one instance, Davis points to shifting emphasis away from traditional intelligence sources to using non-intelligence sources for background analysis (e.g., open source literature). Cultural influences are one aspect of tasking that go beyond the impact of cognitive biases ascribed to individual experts and instead suggest a type of "collective" bias associated with experts' embeddedness within an organization or community.

---

[81] http://www.wsu.edu:8001/vcwsu/commons/topics/culture/culture-definition.html. Accessed on 15 May 2006

[82] INT's (abbreviation for "intelligence domains" such as HUMINT – Human Intelligence, SIGINT – Signals Intelligence)

[83] http://www.gpo.gov/congress/house/intel/ic21/ic21_toc.html Accessed on 16 may 2006.

3.4 Signaling Organizational Expertise

While much of the literature on expert behavior centers on the variability of experts' performance across problem domains and tasks; there is a behavioral constant: *experts signal their expertise*. Experts exhibit behaviors consistent with making explicit there skill areas. Goffman (1959) describes this as self-presentation or building a public image while Becker (1982) views this from the perspective of reputation building as social process. While expertise is largely domain specific, methods of *conveying* expertise may generalize across domains. For example, Jones, (2003) studied architectural firms where legitimacy was established by credentialing their expertise and by embedding the firm within a client network; analogously, a key hypothesis here is that experts signal capabilities much like firms do. Experts may advertise their expertise through artifacts produced, honorifics, roles, and, by embedding themselves within particular work contexts, they establish reputation and build trust. This suggests that the problem of expertise detection may be viewed from the perspective of expertise signaling.

Signaling may be viewed as an incomplete (i.e., asymmetric) information game, Akerlof (1970), where it is assumed that the signaler's "type" is <u>unknown</u> to the receiver. For example, assume an actor is presented as a candidate for a task requiring expertise in information retrieval (IR). The actor *signals* her qualifications (i.e., type) by citing her position as a member of a prominent IR conference program committee. However, only signaler knows her true type — this is private knowledge; the receiver perceives signaler's quality based on how reliable program committee membership is as an indicator of expertise.[84] More generally, receiver's perception of signaler's qualities may be based on signal type, semantics (e.g., message content), and context, so that here receiver may factor in conference affiliation (e.g., ACM[85]), accepted papers, and the reputation or quality of other committee members if known. Based on signal efficacy, receiver may transfer some resource to signaler; for example, a project manager may offer signaler a role on a key task or offer to pay for signaler's expert consulting.

---

[84] Here, behaviors are viewed strictly from a signaling perspective and build in the assumption that signaling behavior is "intentional". It is recognized that this assumption has cultural implications and in some cases may be relaxed somewhat.

[85] Association of Community Machine (ACM)

In this simple game, each player acts to optimize some kind of payoff. This can be viewed as evolutionary design where both signaler and receiver "cooperate" to develop a signaling system of sorts that optimizes costs and payoffs to each. From the example, above, signaler advertises IR expertise using conference committee membership based on prior knowledge as to its usefulness for conveying status. Conversely, if the signal is not effective in conveying expertise (i.e., it is not "honest") then receiver will not reinforce its use. More generally, signal efficacy may evolve over time and across diverse IR groups or settings leading to a type of signaler-receiver coevolution. This suggests that diverse organizations are potentially complex signaling environments in which various signaling strategies arise, die off or, adapt to changes in players and organizational settings.

This view of signaling fits in with Maynard Smith and Harper (2003) who define a *signal* as "any act or structure which alters the behaviour of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved[86]." This definition reinforces the notion that signals go beyond simple "message passing" to fold in adaptive behavior between signalers and receivers over time. Essentially, in this asymmetric information game, the receiver is asking: does the signaler have expertise in information retrieval based on program committee membership; while the signaler harbors the true expertise level. This question is posed with knowledge that the signal may have "cost" and that the signal is typically more costly to generate for a non-expert in IR than an expert.

Costly signaling (as it relates to signal honesty) has origins in animal behavior, Zahavi (1975), and human social status (e.g., wealth), Veblin (1899). According to costly signaling, also known as the Handicap Principle, *signals are reliable because they are costly*. Signals that are easily faked and produced dishonestly are not reliable indicators of the communicated trait. If the signal can not be easily faked (i.e., the cost is too high for non-experts) then it is likely to be reliable. There are two notions of cost here; *efficacy* cost Guilford and Dawkins (1991) which is

---

[86] It is worth distinguishing signals from cues. A signal action has intent and is intended to communicate information about the signaler and to influence the receiver. Signals may be distinguished from an unintentional *cue* such as an inherent characteristic (e.g., weight or eye color). For example, a mosquito flying upwind detects CO_2 from a potential victim; here, CO_2 is a cue but not a signal since the upwind animal does not wished to be bitten, Maynard Smith and Harper, (2003).

the cost associated with ensuring that information is accurately perceived and *strategic* cost which is the cost necessary to guarantee honesty. As such, signal cost does not necessarily imply that the signal is a handicap.   In addition, costs may have various sources. While many signals are intrinsically costly to produce, for example, building *monumental architecture*, Neiman (1998), others are cheap to produce but carry costs through associated consequences, Bliege Bird and Smith (2005). For example, lying (dishonest signaling) may be relatively cheap; however, the repercussions if caught could be significant in terms of lost trust, reduced status, and even legal action. From an expert finding perspective, a ListServ poster may feign (or exaggerate) knowledge of some particular issue by posting on a particular topic. Here, the posts may be cheap to produce; however, the dishonest signaler risks disclosure as to true expertise level, possibly through continuing discussions, requests for help, or follow-up tasking. As such, the consequence for cheap signaling is a type of exposure cost.

While the costly signaling model is generally applicable to a wide range of animal and human social settings, there are alternative models that address signaling conditions in which cost-free signaling may be viable[87].   However, the connection between signaler quality and costly signaling seems to hold in most cases or at least provides a useful starting point for the expertise detection model developed in this thesis.

While there is an extensive literature on animal signaling; for example, see Zahavi and Zahavi (1997), signaling theory has also been widely applied in human social contexts. For example, Sosis and Bressler (2003) studied religious rituals where participation is associated with commitment. Williamson and Wright (1994) examined wealth accumulation as a signal of the ability to produce high quality products. Bloch et al (1999) investigated wedding celebrations as an indicator of social status in rural India. Essentially wedding size or expenditures "signal the quality of the groom's family and thus the enhanced social status of the bride's family." Gambetta and Hamil (2005) argue that signaling is built into every trust game. In a rather

---

[87] There is extensive research in cost-free signaling; for example, see Bergstrom and Lachman (1998) and Lachmann, Bergstron, and Számadó (2000). This includes the case where "signaling can be cost-free when there is no gain in misrepresenting one's condition to anyone." The implications of this on this thesis are discussed as part of recommendations for future research, Chapter 11.

interesting ethnographic study of taxi cab drivers in Belfast and New York, they suggest drivers essentially conduct a behavioral analysis on-the-fly with each new passenger. Based on rider characteristics, the driver's personal preferences and attributes, and the overall setting, the driver determines what degree of trust to ascribe to the rider. For example, drivers may prefer older passengers to younger ones, and wealthier over poorer. In some cases, passenger behavior is viewed as more or less riskier depending on the setting. For example, the authors note that on a Saturday evening in Belfast drivers might expect passengers to be drunk more often than not; on a different night the preference may differ. The overall "decision" to trust depending on particular signals is actually quite complex and, as in the theory of costly signaling, drivers view signals as more reliable if they are not easy to fake and if there are multiple confirming signs. Although the study is somewhat narrow in scope it likely has application in other social (public) settings.

More closely associated with expert finding research are various studies that explore signaling within employment settings. Signaling theory has been used to explain individual behavior in job assignments. For example, Harbaugh (2003) used signaling to explain worker risk taking behavior. According to prospect theory, workers rationalize that low risk gambles (i.e., assignments) that are unsuccessful signal incompetence, while success on high-risk gambles signals strong ability. Jagdish (2004) studied how high-ability managers signaled their abilities through job turnover. He argues that team production masks individual managerial skills so that taking on new roles (through job turnover) across diverse operations reflects individual capability. Promotion has also been shown to be an effective signaling mechanism in various environments; especially in large firms Devaro and Waldman (2005).

Albrecht and van Ours (2006) demonstrate that employers use education to make hiring decisions. They found that the signal value of education increases as the amount of information known about a prospective employee decreases. This aligns with the use of education, to include degree, granting institution, grade point average, and awards, as a basis for hiring new graduates who have no prior work experience. Backes-Gellner and Arndt (2004) examined the role of education as a signal of innovation in a study of start-up companies. They show that education, in the credit and labor market, was even more important for signaling potential success for

innovative start-ups than for traditional start-ups. In addition, completing degree requirements quickly was more important in assessing potential success for entrepreneurs than for traditional new starts.

Signaling theory has been used extensively when individual qualities are not directly perceivable. Signals are visible indicators of hidden innate characteristics such as emotions, intentions, and, potentially, expertise. However, while there are strong parallels to signaling models of animal behavior, and human social settings related to advertising, marketing, status conferral, and others areas, application to expert finding has not received much attention. With that, the hypothesis here is that signaling theory supports an overall framework for the design of expert finder systems.

4 Activity Space Model

Chapter 3 outlined the basic motivation for an expertise search capability based on the notion of expert signaling. The underling premise is that experts signal their qualifications through specific *activities* and artifacts within some organizational setting. As such, the central unit of analysis is the *activity space* (AS); a sampling frame of sorts that binds expert signaling behavior to a particular work context. The remainder of this chapter lays out key elements of the AS framework.

4.1   Activity Space Concepts

Activity Spaces, viewed broadly, are used across multiple disciplines and problem domains such as geography, urban planning, and anthropology. As such, there is considerable variation in terms of purpose and structure. For example, in zoology, an activity space may be defined as the "range or 'spectrum' of environmental conditions and habitat characteristics that support the normal activity of an organism", Rickleffs (1990). Kopec (1995) characterizes functional (physical and mental) impairments in terms of restrictions on a person's activity space ("a multidimensional space that represents human potential for activity"). In computer systems an activity space "groups multiple task-specific actions into a logical set and provides the programmer with base functionality."[88] Similarly, SEPIA[89], a system supporting multi-user authoring, uses activity spaces as the central analysis unit consisting of four tasks: planning, representation and structuring, development and representation of argumentative structures, and document organization for the target audience.

In this research, an activity space is an information space populated by actors performing actions using tools and artifacts consistent with a goal and constrained by rules specific to that space.

[88]http://e-docs.bea.com/alui/devdoc/docs60/Customizing_the_Portal_UI/Creating_Custom_Activity_Spaces/plumtreedevdoc_customizing_activityspace_intro.htm Accessed on Mar 16, 2006.

[89] http://www.gmd.darmstadt.de/concert/activities/past/sepia.html Accessed March 18, 2006

For example, a ListServ may be viewed as an AS. A goal or outcome of a ListServ user may be to exchange or disseminate information across a community. ListServs have membership and work is accomplished through message posting on particular topics. Actors use tools to support goal attainment and this may include email capabilities as well as models, concepts, and theories used to frame discussion consistent with the objective. Actions are restricted (i.e., there are "rules") consistent with the ListServ environment to include topic restrictions dictated by ListServ "owners", privacy and intellectual property controls, and membership limits. The AS reflects individual actor behavior (e.g., message posting) as well as "community" interactions as evidenced through threaded discussions tied to a particular topic. While most ListServ discussions are peer-to-peer in nature there may be a division of labor; i.e., "roles" carried out within a thread or across multiple themes that lead to asymmetric relationships amongst members. These roles may be based on self-organization (e.g., actors assuming balancing positions on controversial topics) or the roles may be dictated by the formal organization or work assignments.

While the ListServ example and others may serve to introduce AS elements, a more formal underpinning is provided by Activity Theory. Activity Theory (AT) has its origins in early 20th century Russian psychology. A history of AT may be found in Leont'ev (1974), Kaptelinin (1996), Kuutti (1996) and Nardi (1996). AT is not strictly speaking a "theory" but is more a conceptual framework providing basic principles for which to understand work practices. The framework provides a way of analyzing actions and interactions within a particular context. The central unit of analysis is the *activity*. Activities consist of actions or sequences of actions related to the activity goal and motive. The notion here is that actions cannot be interpreted, "without a frame of reference created by the corresponding activity", Kuutti (1996). Activities have the following characteristics:

- *Activities* consist of specific actions or action sequences performed by a subject (actor) and focused on an objective.
- *Actions* are carried out within an activity and are guided by goals although different actions may be used to accomplish a goal. As examples, an actor may publish a document with the goal of communicating recent research findings or in a different activity an actor may publish a document to share project financial data.

- *Operations* are lower-level procedures used to perform actions. For example, "reply" is an operation performed when responding to an email.

Figure 4-1, shows the three levels with two examples. The first example is from Kuutti (1996). The second maps more directly to expertise signaling. Here, the activity has the motive of conveying expertise—an overt case where an expert intends to signal skills. The activity is performed through various actions carried out; for example, for a ListServ activity, actors register for a (topic-specific) ListServ, and post on the list topic. Each action is associated with a goal; for example, an actor registers in order to become a ListServ member so that he may post or have access to ListServ postings. Operations may consist of using the ListServ software to communicate; for example, using "New Message", "Reply" or "Delete" options.   The levels are fluid and, for example, actions in one context may become operations when the goal changes; similarly, activities may become actions.

Activity Level: Motive	Carrying out research on a topic	Conveying Expertise on a Topic (Signaling)
Action Level: Goal	Searching for references; Participating in a conference; Writing a Report	Registering for a ListServ in a particular topic area; Posting a topic-relevant message;
Operation Level: Conditions	Using logical syllogisms; Selecting appropriate wording	Using ListServ operations: "New Message", "Reply" or "Delete".

Figure 4-1: Hierarchical Levels in an Activity (Adapted from Kuuttti (1996))

The *activity system*, Cole and Engestrom (1991), provides a visual depiction of all elements in an activity; Figure 4-2.   Here the subject refers to the actor or group whose point of view is taken. For example, using the ListServ example, actions performed within a particular activity may be viewed from the perspective of a specific poster or from a group of posters participating within a communication thread. The perspective taken depends on the purpose of the analysis. Each activity has a motive; for example, to alert target groups to an actor's expertise in nanotechnology. The objective, then, may be to post messages about a new nanotech design method. Instruments are internal or external mediating artifacts that transform the objective into

an outcome. Artifacts are used to shape activity (e.g., use of a search engine to locate information) but they are also the product of activity; i.e., a retrieval list.

The community is made up of other actors who share the objective with the subject; this may be all ListServ members, a project team or meeting attendees. Rules shape actions taken by the subject and interactions within the community. In a ListServ, rules may dictate the range of sub-themes accepted within a ListServ—based on charter, as well as explicit and implicit discourse norms. Some rules may be formally mandated others built up informally from group consensus. The division of labor determines how the activity is distributed across the community. From an expert ranking perspective, role attribution (division of labor) may be used to confer expertise status. While activity theory provides a rich conceptual framework for, say, designing collaborative spaces or analyzing extant work contexts, it is used here to inform the design of an Activity Space (AS), useful for expertise detection.
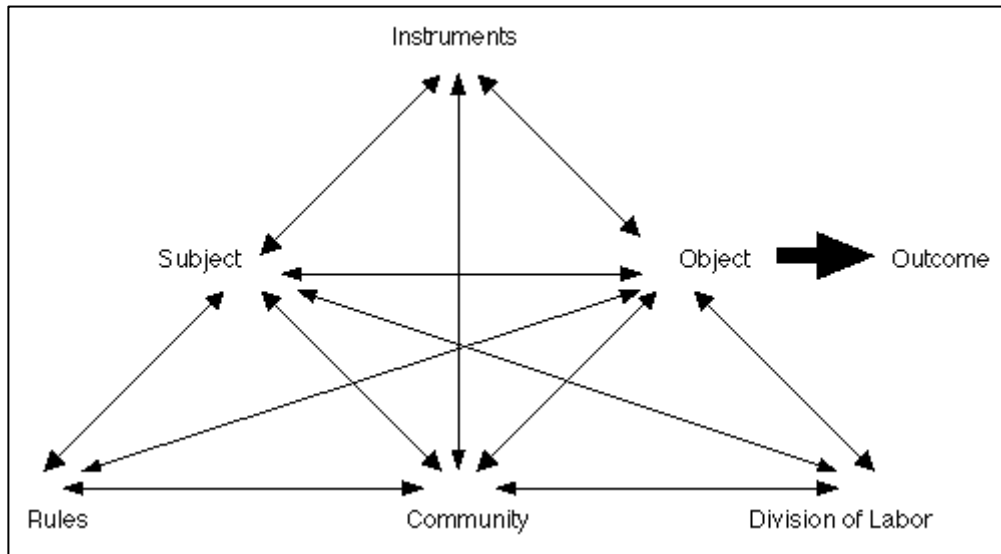
Figure 4-2: Activity System

4.2 The AT to AS Translation

The discussion so far paints only a cursory view of AT; a more detailed exploration of the key concepts and their application in various disciplines to design and analysis are outside the scope

of the paper. However, this surface view of AT provides a foundation for developing the AS framework for use in the *Expert Locator* model. Essentially, AT is used to inform AS.

There is dual motivation to ground AS in an activity theory framework. First, as noted above, activity theory provides a rich conceptual space from which to address context and specific work elements such as the notion of actors, community, and mediated actions. Second, there is shared perspective on the relationship between "actions" and "expertise signaling".   In the AS-based model, experts signal their skills and experience through actions and use of artifacts (e.g., authoring a paper). In activity theory, "you are what you do", Nardi (1996). This link between *actions* within an activity and *signals* within an activity space is a foundation of the *Expert Locator* model discussed in Chapter 6.   In the remainder of this section, an activity space framework is developed reusing key elements of activity theory.   The intent here is to retain selected elements of AT as part of an AS framework

The activity space schema is shown in Figure 4-3 as a collection of elements and their relations. Each AS element will be discussed here with references to like AT components. In the activity space, <u>Actor</u> is a primary element. There are four primary relations in the model: *membership*, *co-actor*, *actions* and *association*.

- Actors are affiliated with a particular AS; they have <u>membership</u>. Membership may be conferred through formal registration, based on activity, roles, or other bases.
- Actors are linked to other actors based on membership or interactions; as such, co-actors may reflect total membership or they may be specific subgroups organized around joint actions. For example groups may form around ListServ discussion threads[90].
- Actors perform <u>actions</u> consistent with space type. Some actions may be AS-specific. For example, Meeting actions may include "schedule", "invite", "cancel", "accept", and "reject". Other actions, e.g., "query", may be more general and applicable to multiple AS

---

[90] Structural relationships between actors may also be imposed on the AS by external applications such as the *Expert Locator*. More generally, the AS is a data object that may be operated on by various functions. This provides a useful decoupling between AS data objects and external applications.

contexts. There is a rough equivalence between *actor,* as used here, and *subject* in the AT framework, the AS *co-actor* maps to the AT c*ommunity*. A*ctions* are used in both AS and AT.

- Actors are <u>associated </u>with *artifacts*. For example, an actor may be associated with a use of a particular tool or document authorship.

There are several AT components not addressed explicitly in the AS model. For example, division-of-labor, which addresses work distribution across the community, is implicit in the membership relation; where model implication allows for a degree-of-membership qualifier (i.e., weight) on the membership relation. This may be based, for example, on an "effort" model which is used to estimate division of labor using actual labor usage reported in a project. Rules are also not explicit in the current AS framework; however rules constrain behavior and this may bias the kinds of actions performed as well as the community structure that evolves.
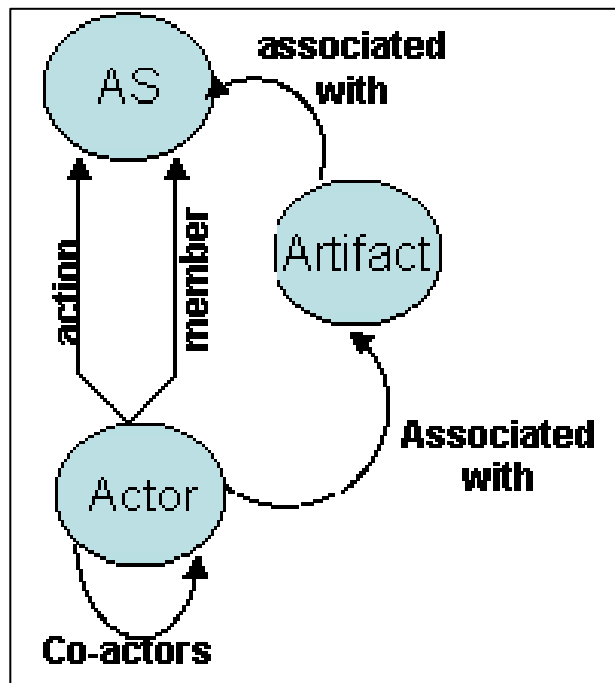
Figure 4-3: Activity Space Schema

Figure 4-4, below, contrasts two AS's: ListServ and Meeting, selected for illustrative purposes only. Actual selection of an AS for use in the *Expert Locator* model is left to Chapters 5 and 6.

98

In the Meeting AS, meeting attendees (actors) perform actions such as scheduling the meeting and inviting participants as well as tasks related to performing work. Artifacts may consist of specific tools used at the meeting (e.g., whiteboard) or may be actual work products such as a workshop report. The ListServ AS can be viewed similarly.
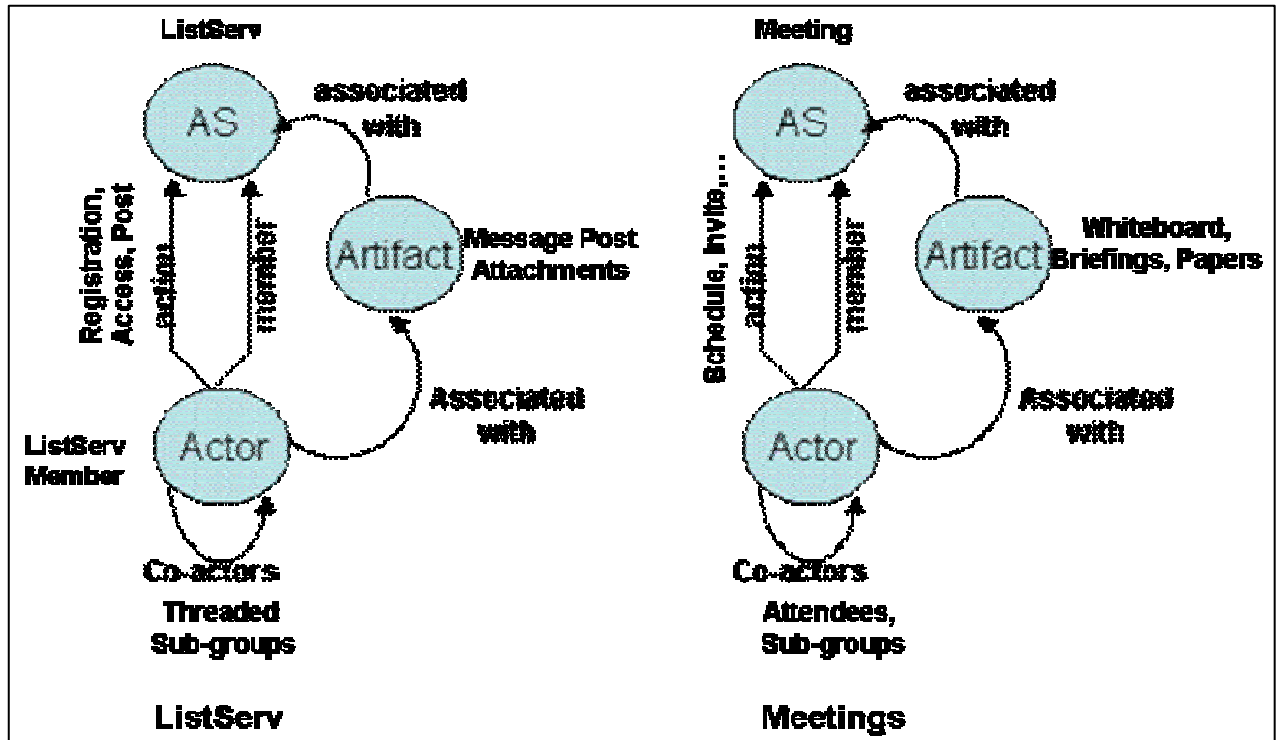
Figure 4-4: Activity Space Schemas: ListServ and Meeting

4.2.2 Attributes

As part of the overall schema, attributes are associated with each AS element. This is a departure from activity theory which does not directly incorporate characteristics of actors, actions, operations, and artifacts. Actors have organizational attributes such as *name*; *hire date*; *home department*; *job function*; *rank*; and *personal contact* information. Artifacts (e.g., documents) have metadata such as *authorship*, *title, generation date*, and *genre*. Instruments or tools may have other metadata such as *tool type* (e.g. visualization tool). Activity space attributes include *title*, *genre, origination date*, and other descriptors.

Figure 4-5 depicts the MATLAB activity space which is a Corporate Technical Team organized as a Cluster Group (a loosely organized but formally assigned group working asynchronously). The group formed in February 2000 and focuses on MATLAB applications. John Smith is a MATLAB member who is a simulation engineer assigned to Department G060; his office location code can be mapped to McLean,, Virginia. John Smith is an author of a whitepaper "MATLAB Simulation" written in October 2004. Attributes provide further "context" to actions and relations associated with the AS.
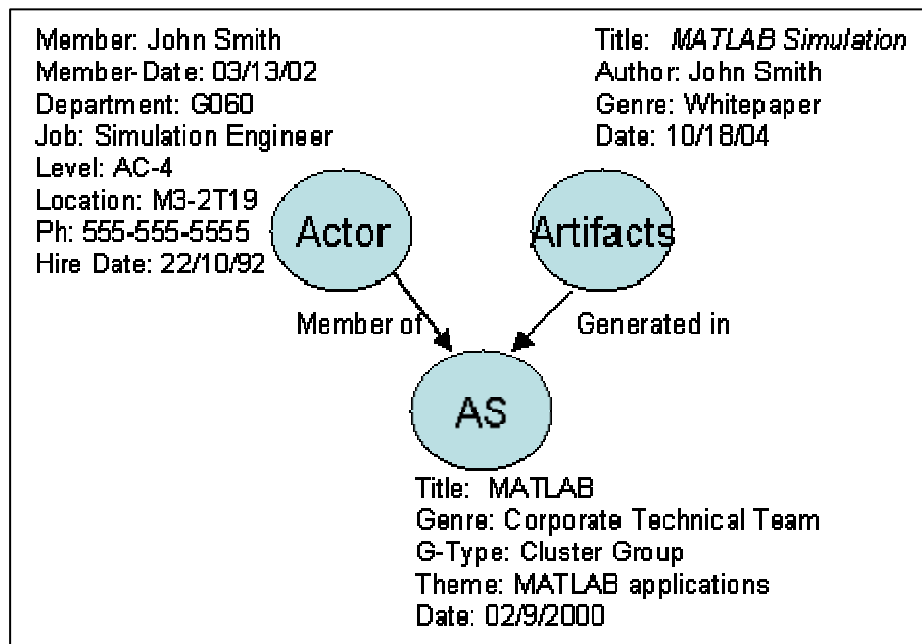
Figure 4-5: Attribute Types (MATLAB Cluster Group)

Attributes can be used to discriminate amongst individual actors within an expertise network based on organization ties, physical location, or other characteristics. In the following example, actors with expertise in enterprise architecture have been selected from a larger group of enterprise architects based on participation in a series of four technical exchange meetings. Each meeting is an activity space (AS-1, AS-2, AS-3, and AS-4) as shown in Figure 4-7. The actor nodes are sized according to the number of meetings attended; that is, node size relates to the number of "attendance" *actions* across the four activity spaces. Color coding actor nodes by site (home base) provides some indication that the actors that attended the most meetings are most often from a particular site; here color coded green. This analysis can be used as part of more

general post-retrieval analysis to support selection (choosing experts to contact) or to identify actors playing particular roles for example.
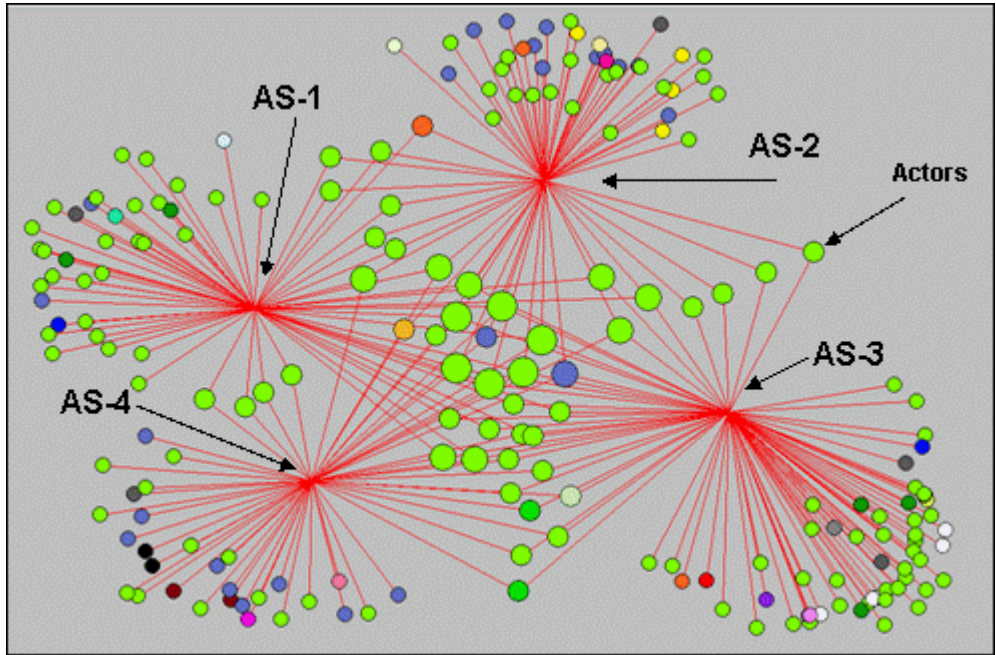
Figure 4-6:   Actors Distributed Across Activity Spaces

4.3   Additional Activity Space Perspectives

4.3.1   The Dual of the AS Model

The AS model is activity centric; given its roots in activity theory. The AS captures collective expertise from multiple actors within a particular setting. The dual of the AS is ego-centric and focuses on actor expertise as distributed across multiple activity spaces. The complementary nature of these models allows for viewing activity spaces as populated by experts and conversely for viewing experts in terms of the activity spaces they are embedded within. The relationship between the ego- and AS- centric models is reflected in the 2-way table (Table 4-1) where rows are assigned to actors and columns to AS's. Row margins characterize the summation of individual evidence across activity spaces, while column margins reflect the aggregate evidence from multiple experts within a particular activity space.   The main point here is that in the

101

aggregate, analysis of evidence distributed across activity spaces is equivalent to assessing the distribution of evidence across experts.

	AS-1	AS-2	...	...	AS-m	
Actor-1						Σ
Actor-2						Σ
Actor-3						Σ
:						
Actor-n						Σ
	Σ	Σ			Σ	

Table 4-1: AS-Actor, 2-Way Table

4.3.2   Activity Spaces and Personal Networks

The AS-Actor view, Table 4-1, forms an affiliation network, Wasserman and Faust (1994), where each actor is described in terms of his/her membership within specific activity spaces. Therefore the two-mode actor-activity space graph can be transposed into two single mode graphs; AS-AS and Actor-Actor. Here, the actor-actor graph can be used to identify each actor's personal network, where, each personal network is generated by taking each actor as the central node (ego) linked to actors co-located with ego in one or more activity spaces and who share expertise in some domain.   While a personal network is defined "liberally" here since co-work may not satisfy all criteria for inclusion in an actual personal network; in this context it reflects the actor's likely "awareness" of alters if not true collaborators.   Nardi, Whittaker, and Schwarz (2002) view intensional (personal) networks as central to a wide range of work practice and maintain that "the most fundamental unit of analysis for computer-supported cooperative work is *not* at the group level for many tasks and settings, but at the individual level as personal social networks come to be more and more important." While personal networks as defined here may

not match fully the notion of intensional networks[91] constructed by the ego, they do suggest which actors have correlated work profiles with overlapping activities; as such they may approximate actual intensional networks.

As discussed in Chapter 6 and 7, the *Expert Locator* model implicitly accounts for personal networks through measures of social context. Figure 4-7, below shows a personal network for a particular topic built. While the personal network provides a basis for assessing ego's expertise level it can also be used to support *selection—identifying an actor to contact*. In particular, a system user can use knowledge of ego's personal network to identify referral chains, Kautz et al (1997), or to identify contact surrogates. In many cases users may not wish to contact the expert directly and here the personal network may be useful for finding those with like expertise.
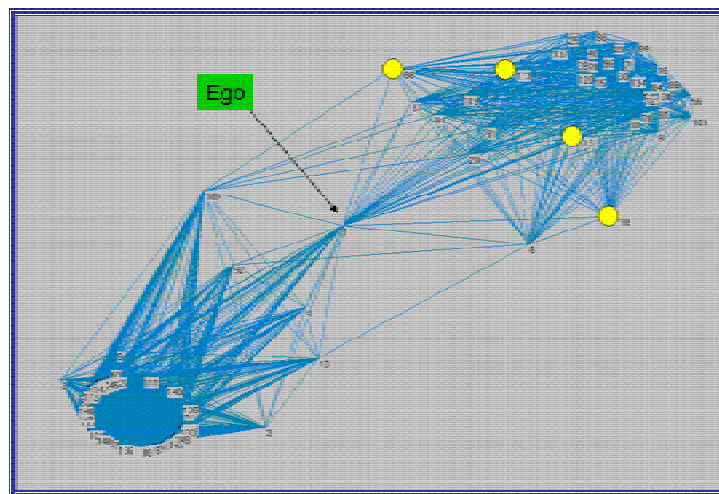
Figure 4-7: A Personal Network with Ego in a Bridge Position between Two Groups

In the next chapter, operational work contexts are identified, assessed, and selected for use in the *Expert Locator* prototype.

---

[91] Intensional networks reflect an ego's <u>deliberate</u> effort to <u>construct</u> and <u>manage</u> their social network.   From an AS perspective, personal networks may vary as to their "intensional" nature.

5 Enterprise Activity Spaces

In this chapter, enterprise activity spaces are identified, categorized, and selected for use in the *Expert Locator* prototype. Activity Space *integration* into *Expert Locator* prototype is covered more fully in Chapters 6 and 7.

5.1   Activity Space Taxonomy

MITRE has a number of diverse business forums that, modeled as activity spaces (AS), are integral to day-to-day business. Collectively, they represent widely varying work contexts that cross-cut formal and informal work as well as individual and group activity. A survey of the target environment resulted in sixteen spaces organized here into four AS classes; Table 5-1. The spaces identified below, are more fully described in Appendix C.

Activity Space Class	MII: Activity Spaces	Expertise Aspect
Organization/Personal	Public Share, Private Share, Blogs, About-Me, E-mail, Instant Messaging (IM)	Personal spaces used to convey user interests, knowledge, or expertise. Each personal space is linked to user's home organization (e.g., department).
Corporate Technical Teams	Technology Area Teams (TATs), Skills Clusters, The Hotline, MITRE Repository of Knowledge (MRoK)	Team-based spaces formed around corporate teams and related to specific expertise areas or expertise services
Projects	Project Page, Project Share, SourceForge	Team-based workspaces set up to organize, store, and share project work consistent with access constraints (e.g., privacy or security)
Community	Sharepoint, ListServs, Technical Exchange Meetings (TEMs)	Collaborative spaces that support multi-user communication and information sharing.

Table 5-1: Activity Space Classes and Instances within the MII

1. *Organization/Personal* spaces capture individual behavior in the context of one's organization home (either at the home department or corporate level). Work of this type is not easily traced to projects (i.e., it is <u>not</u> generally associated with a project charge

number) but it can be tracked to an individual and linked to home organization. A brief description of each activity space follows:

- o *Public Share:* To promote knowledge sharing, each employee and contractor (with system access) has a Public Share folder. Public Share folders are typically used as a type of personal information space. Users can drag-and-drop documents into their folders for sharing and at the same time publish documents to the corporate collection. A Public Share folder can be hierarchically organized into subfolders.

- o *Private Share:* Private Share folders are structurally equivalent to Public Share folders except that Private folders are accessible only to their owners, and require MITRE domain authentication. Users have read/write access only to their own Private folder.

- o *Blogs:* Blogs @ MITRE is an interactive content management system that provides a simple way for all MITRE employees and contractors to post information regarding their individual or project work.

- o *About-Me:* The About-Me folder can be used to publish professional information about a user's skills and experiences. The About-Me folders can be written to only by their owners, and require MITRE domain authentication. Other users can view files in an About-Me folder as well as copy files from a person's About Me folder. About-Me is semi-structured; users may use "fields" or "tags" to denote certain entries in their description.

- o *Email:* Microsoft Outlook is the primary email system used at MITRE. It offers integrated mail and calendar features. Users may use the main email client or access email using Outlook Web Access. Analysis of email message text, for the purpose of creating awareness of user's interests and skill areas has been explored in experiments[92] that preceded this thesis.

- o *Instant Messaging:* AOL Instant Messaging (AIM) is a messaging system used by many MITRE staff. AIM is often used in situations where a phone call to a colleague or to someone on a support team is not feasible. AIM is neither

---

[92] Early MITRE-internal analyses were performed on volunteered email text and header information and used to compare with a commercial product (ActiveNet) that exploited email for expert finding.

supported nor endorsed by the MITRE's Information Security Committee but works through the corporate firewall. Since AIM has a weak security model MITRE users are advised to not discuss sensitive topics; it is not difficult for an intruder to masquerade as the desired recipient or sender of AIM messages.

2. ***Corporate Technical Teams*** capture behavior associated with formal groups assigned to a technology or business area. Technical teams may provide technology assessments, steer corporate research, or provide business area assessments. A brief description of each activity space follows:

   o *Technology Area Teams (TATS):* Technology Area Teams (TATs) are part of MITRE's Technology Program (MTP) directed by MITRE's Chief Technology Officer; and supported by Chief Engineers from each operational center. Each TAT consists of technical experts from across MITRE's operational centers. TATs prepare forward looking assessments on current and emerging technologies, support proposal review during the MTP research funding competition, and generally provide support to staff members and sponsors in areas related to their expertise.

   o *Skills Clusters:* The objective of Skill Cluster Groups is to keep MITRE personnel abreast of technology developments. Cluster Groups are organized around various special interests and skill areas and are committed to disseminating technical information and providing referrals to outside and internal experts.

   o *The Hotline:* MITRE's Technical *Hotline* is an on-line service providing staff access to resident experts in a number of technical areas. The *Hotline* service uses a peer-reviewed registration process for assigning experts into topic areas. Typically 3 to 5 experts are in each category. Users can email questions to *Hotline* experts using an online form that allows users to enter questions and link questions to one of the 33 established expertise areas or to the "Other" category. The appropriate expert answers the question and archives the question and answer for searching and analysis of question trends.

   o *MITRE Repository of Knowledge (MRoK):* MRoK is a knowledge management initiative focused on capturing knowledge directly from MITRE staff. Users post questions and answers to topic categories. MRoK has no formal registration in

terms of an established cadre of experts; instead experts are *attracted* to questions within one or more domains. The lack of formally recognized experts distinguishes MRoK from Hotline. The system, however, has similarities with ListServs, as described below, in that domains are established and threaded discussions are possible.

3. ***Projects*** reflect formal work, both internal and sponsor-funded. Projects capture formal tasking that subsumes the bulk of work performed; projects are typically partitioned into subtasks which are organizationally tracked in terms of labor, staff assignments, and artifacts produced. Project data are split between a Project Page which includes standard project metadata (such as task membership, labor charges, owning organization, and sponsor affiliation) and the Project Share Folder which contains task artifacts or documents. SourceForge, below, represents a special project class. A brief description of each activity space follows:

   o *Project Page:* Each project page includes the task name, a short description (label), the parent project (as most projects have multiple tasks), task leader, and period of performance. This is followed by a list of task members to include their home department and level of effort (percentage of total task labor used). The page links to Project Share which contains all documents archived to the task.

   o *Project Share*: The Project Share Folder system is a Web-based environment for knowledge sharing and reuse. Project Share allows MITRE users to publish and share project-related documents and files. Access to documents and files shared is available by browsing folder hierarchies, or by searching.

   o *Source Forge: The* SourceForge[93] server (iSF) internal to MITRE provides developers with access to a wide range of tools including bug tracking, task management, code versioning (CVS), mailing lists, forums, and project web pages. iSF can provide evidence of software development or application expertise and can be used to identify development teams and link teams to sponsor programs.

4. ***Community*** spaces capture large (often self-organizing) group activities focused on particular problems, technologies, or business areas. Here, groups may form out of

---

[93] http://sourceforge.net/ Accessed on October 15 2005.

mutual self-interest as opposed to corporate mandate. A brief description of each activity space follows:

- o *ListServs*: A ListServ may be viewed as a mechanism for forming communities; a sort of communityware that supports self-organization around selected business or technology issues. ListServs are essentially open forums for dialogue on various topics. MITRE maintains *Corporate Lists* and *Shared User Lists*.
  - Corporate Lists are managed automatically using information obtained from the MII Intranet. They are aligned with MITRE organizational and geographical entities (such as departments, centers, and sites), and for the various MITRE job titles. They include MITRE employees only, and may only be used by MITRE employees. Their purpose is largely administrative.
  - Shared User Lists are created and managed by MITRE employees, and are usually related to a particular MITRE project or topic area; Shared Lists have domain focus and are aligned with expert finding.
- o Technical Exchange Meetings (TEMs): TEMs are internal meetings held by MITRE employees for MITRE employees. Generally, technology experts or business stakeholders organize a TEM to address a compelling technology or business issue. Each TEM is archived to include attendance lists, briefings and papers, and summary findings. A TEM is typically follows a "workshop" format and includes a brief description consisting of the TEM theme, target audience, and registration requirements (if any).
- o *Sharepoint:* Community Share is a pilot project using a community-based document management product, Sharepoint[94], to address MITRE's requirements for team support. Sharepoint is a community- or team- based collaboration platform that provides a common web space for working on shared documents, posting events and announcements, posting links to web sites, having threaded discussions, and tracking action items or agenda items.

---

[94] http://www.microsoft.com/sharepoint/ Accessed on 18 October 2005.

5.2 Activity Space Selection

Enterprise activity spaces organized according to Table 5-1, above, vary in terms of their usefulness for expert finding. During this research, some spaces were relatively mature in terms of implementation and actual usage, others were newly emerging, and some were scheduled for "retirement". Within this shifting operational context, selection criteria were set up to aid in identifying activity spaces which would be viable for use in the *Expert Locator* prototype. The overall methodology and actual selection are discussed below, and supported by activity space descriptions provided in Appendix C. Integration of specific activity spaces into the expertise model and operational prototype is discussed in Chapters 6 and 7 respectively.

5.2.1 Activity Space Selection Criteria

Activity Space selection criteria used here can be viewed from the perspective of the host *environment* or activity space *composition*. Environment factors address needed infrastructure support such as *access* and *policy compliance*. Space composition criteria are more focused on the kinds of signaling evidence (behaviors and artifacts) needed to assess expertise. Collectively, the criteria provide a basis for answering: is an activity space "accessible" and "are expertise signals *visible*?"

AS Criterion	Measurement	Assigned Values	Comment
Expertise Relevance	The level that signaling evidence is relevant to expertise assessments?	**High, Medium, or Low**	Some AS may provide a richer context for gleaning skills and experience than others.
Policy Compliance	Is AS access policy-compliant and aligned with privacy?	**Yes or No**	Formal policy or cultural norms can affect access.
Data Access	Is signaling evidence supported by enterprise services or applications?	**Yes or No**	Infrastructure support has implications for prototype design and development.
Attribution	The level that evidence is attributable to a particular actor, group, or activity?	**High, Medium, or Low**	Various attribution levels are possible; there may be missing co-authors
Artifact Signaling	The level that artifact evidence is associated with the target AS?	**High, Medium, or Low**	Artifacts vary across spaces; e.g., postings and labor *burn* rates.
Social Signaling	The level that social evidence is associated with the target AS?	**High, Medium, or Low**	Includes organizational membership, co-work, co-authorship, etc. .

Table 5-2: Selection Criteria

Table 5-2 describes the six main criteria used. Criterion values are based on type; some are binary; i.e., *Yes* or *No,* while others are ordinal: *High, Medium, or Low*. Assessments are based on organizational knowledge, activity space contents, and usage data where available.

5.2.2   *Activity Space Culling*

Selection criteria are applied to each AS, Table 5-3.   While the assessment is largely qualitative, in selected areas there are empirical results that support the rating levels used here. For example, ListServs are rated "High" on *Expertise Relevance*, *Artifacts*, and *Social Presence*; as supported by data in Appendix C. ListServs have numerous postings covering a range of related topics within some domain, and have diverse membership needed to support rich interaction. Lists are rated "High" on attribution since posting headers and thread tracking supports message attribution on single and multi-posting topics. While most Lists are publicly accessible; List owners may restrict rehosting List content (at a central location) for uses other than standard message dissemination and review by members. This suggests that some Lists may not be incorporated in the *Expert Locator* collection based on privacy concerns; however, this is not common and, therefore, there is not a Policy Compliance issue.

While, each activity space may be assessed on a criterion-by-criterion basis a more direct method for discriminating amongst spaces is to successively apply the most discriminating criterion as a basis for partitioning the candidate set. The selection strategy is characterized in Figure 5-1. A decision tree is set up to partition activity spaces step-wise, top-to-bottom, and left-to-right using highly discriminating criteria. For easy viewing, Figure 5-1 provides a "compressed" decision tree showing two sequential filters leading to the Selected Spaces box[95]. From Figure 5-1, at the "root", all activity spaces are initially assessed. The first filter (testing *Policy Compliance*) rejects spaces that did not pass corporate policy or privacy restrictions (i.e., ***No*** is entered); for example, Private Share owners have access restrictions that preclude read/write access necessary to support expert finding. This precludes using Private Share folders in the expertise model.

---

[95] Note a sequential logic is used here and the most restrictive conditions are addressed first in the sequence; as such some spaces eliminated on one test may also have been eliminated on subsequent tests.

Activity Space Class	Activity Space	Expertise Relevance	Policy Compliance	Data Access	Attribution	Artifacts	Social Presence
Organization /Person	Public Share	High	Yes	Yes	Medium	High	Low
	Private Share	High	**No**	**No**	Medium	High	Low
	Blogs	High	Yes	**No**	High	High	Medium
	About-Me	High	Yes	Yes	High	Low	Low
	e-mail	High	**No**	**No**	High	High	High
	Instant Messaging	Low	**No**	**No**	High	Medium	Medium
Corporate Technical Teams	Technology Area Teams (TATs)	High	Yes	Yes	Medium	Medium	Medium
	Skills Clusters	High	Yes	**No**	High	Low	Low
	The Hotline	High	Yes	**No**	High	Low	Low
	MRoK	High	Yes	**No**	High	Low	Low
Project	Project Page	Medium	Yes	Yes	Medium	Low	High
	Project Share	Medium	Yes	Yes	Medium	Medium	Medium
	SourceForge	High	Yes	**No**	High	Low	High
Community	ListServs	High	Yes	Yes	High	High	High
	TEMs	High	Yes	**No**	High	Medium	High
	Community Share	High	Yes	**No**	Yes	High	High

Table 5-3 Activity Space Assessments

E-mail may have significant potential for inferring expertise. E-mail based social networks (who-emails-whom) can be used to identify work groups, key persons, and organizational ties. Analysis of message text has also been addressed in experiments[96] that preceded this research effort. However, MITRE e-mail is not viewed as a public resource which can be openly shared and analyzed; there are a number of privacy concerns that arise even where owner identity and message content are protected. Corporate policy precludes use of E-mail at this time.

AOL Instant Messaging (AIM) is used by many MITRE staff. AIM is neither supported nor endorsed by the MITRE's Information Security Committee but works through the corporate firewall. Since AIM has a weak security model MITRE users are advised to not discuss sensitive topics, in addition, it is not difficult for an intruder to masquerade as the desired recipient or sender of AIM messages. As with email, privacy concerns preclude near-term use of AIM

---

[96] Early MITRE-internal analyses were performed on volunteered email text and header information and used to compare with a commercial product (ActiveNet) that exploited email for expert finding.

messaging. Using the *Policy Compliance* criterion as a filter, Email, Instant Messaging, and Private Share are eliminated.
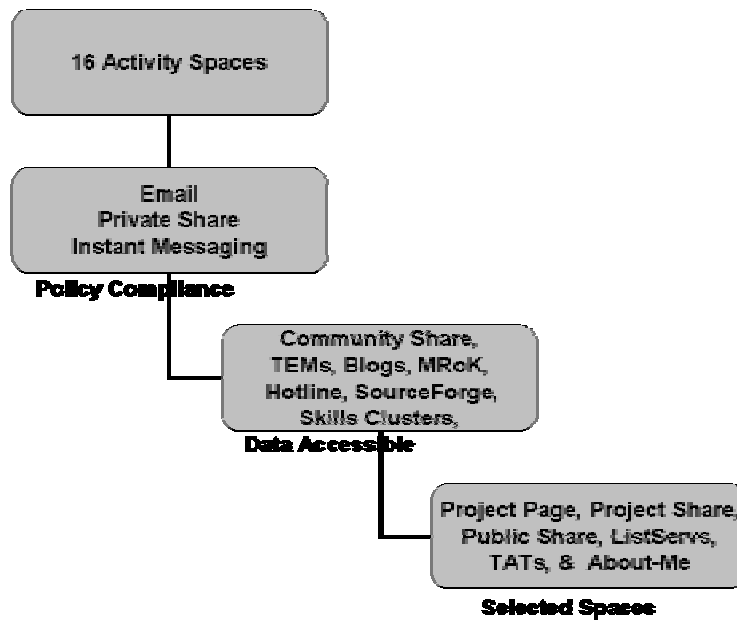
Figure 5-1: Sequential Activity Space Selection

The remaining *Policy Compliant* spaces are then compared to the *Data Access* criterion and, as noted; seven spaces lacked needed infrastructure support or required significant development to provide as part of this effort. Community Share and TEMs are of primary interest here, especially given long-run enterprise direction. While Community Share is expected to subsume several existing spaces in the future, and to provide needed infrastructure for a wide range of formal and informal activities, it is not well supported currently. TEMs are a rich information resource as already discussed; however they are not well organized, not indexed for retrieval, and the data are largely inaccessible since much of it is distributed across personal desktops not accessible from the network. Similarly, Blogs, MRoK, The Hotline, SourceForge, and Skills Clusters were not indexed by the corporate search engine at the time of this research and were therefore not considered for the initial prototype. All seven spaces are eliminated from further consideration. With that, the selected spaces (last box) satisfy corporate policy restrictions, and have no data access constraints that cannot be managed by using enterprise services or low-cost, custom applications.

5.3 Selection Refinement

In the previous section, a preliminary selection set was culled out based, largely, on two main criteria: access policy compliance and supporting infrastructure (i.e., access to services or applications). However, some spaces are potentially more viable than others and there is some basis for consolidation. For example, About-Me, has low utilization but is currently indexed by the corporate search engine making the data readily accessible. As such, it is reasonable to consolidate About-Me with Public Share into a combined AS. Here, merging addresses potential data sparseness; acting as a "smoothing" operation. In addition, since TATs may be viewed as a special internal project as well as a Corporate Technology Team, it is reasonable to move it to the project space; this also should reduce sparseness resulting from using TATs as a single AS.

This suggests the following activity space definitions, Table 5-4. **Project** space now subsumes the Project Page (membership, roles, and labor usage), Project Share (project artifacts) and TATs. **ListServs** are preserved; there is no combination with other spaces or transformation specified. The **Organization/Person** space includes Public Share, and About-Me.

Selected Prototype Activity Spaces	Corporate Space
Project	Project Page, Project Share, TATs
ListServs	ListServs
Organization/Person	Public Share, About-Me

Table 5-4: Prototype Activity Spaces Built Mapped to Corporate Spaces

This collapsing, as a general strategy, leads to coarser partitioning of the enterprise workspace. However, there is certain utility in being able to incorporate spaces into the expertise model directly or as part of some *super-space* aggregate; since aggregation, for example, may lead to more stable expertise rank estimates.

As discussed, this chapter is focused on selection, AS *integration* into the expertise model is addressed in Chapters 6 and 7.

6 Formal Expertise Model

This chapter develops an expertise model enabled by signaling theory and activity theory discussed previously in Chapters 3 and 4, respectively. While this chapter integrates signaling concepts into the formal expertise model, the reader is referred to Chapter 7 where model implementation is addressed more fully; to include a description of various sources used to instantiate the working prototype. Finally, theory meets practice in Chapter 9 where experiment results are viewed in the context of Signaling and Activity Theory.

6.1 Unified Framework for Expert Finding

Signaling theory provides a rich framework in which to explore expertise detection. A signal is *an act or structure that alters the behaviour of another organism, which evolved because of that effect, and which is effective because the receiver's response has also evolved*, Smith and Harper (2003). Relevant to expertise detection, a signal may be a publication that influences a target audience, so that publication and audience response co-evolve to reinforce each other. Similarly, participation in TREC may provide researchers notoriety within the IR community since it signals researcher knowledge and skills. Clearly, there is an inherent *cost* associated with TREC participation that goes beyond project labor expenditures and more directly relates to skills needed, results quality, collaboration with peers, and links to prior work. In most cases, TREC participation by non-IR specialists would be cost prohibitive. It would be difficult to mimic what the IR expert knows and accomplishes so that TREC participation is a reliable signal of IR expertise. This falls in line with costly signaling theory and the handicap principle, Zahavi (1975), as discussed in Chapter 3. Signaling theory motivates expertise models that associate signal cost to expertise. The working hypothesis is that costly signals are more indicative of expertise than minimal cost signals that may be easily produced by an entire population; not just those possessing certain expertise.

From an enterprise perspective, signaling captures interactions between multiple signalers and receivers; reflecting varying organizational contexts, activities and social interaction. However, expertise exchange in heterogeneous work settings leads to signaling models that are generally complex and unwieldy to formulate. In addition, there may be a wide range of signal sources and types that complicate discerning signal reliability. This can be exacerbated by deception as signalers may either exaggerate qualities or hide their traits from others who may otherwise wish to make their expertise visible. This inherent complexity suggests a simpler approach; at least at this stage of research.

Figure 6-1 (a), depicts a 2-person signaling model. In this simple, asymmetric information game, *signaler (S) signals receiver (R)*, within some organizational context. In this model signaler and receiver adapt so that signal design is optimized to reflect signaler and receiver payoffs; signals evolve in a way that benefits both signaler and receiver. As a matter of completeness, receiver transfers some resource to signaler (e.g., confer status or reputation, award tasking). While this model is somewhat simplified it is potentially quite complex as it allows for complex receiver behavior (i.e., resource transfer), signal adaptation based on signal-receiver co-evolution, and the case where actors are both signalers and receivers.

Signaling theory, as used here, is less ambitious as suggested in Figure 6-1 (b). Here, the focus is reduced to characterizing *signaler*, *signal*, and *context*. To incorporate *Expert Locator* into the model, the *expertise model* is inserted as an adjunct to receiver; acting "passively" to analyze signaling evidence. While the receiver may view signals directly; the expertise model serves to rank order candidate experts based on aggregate signaling evidence related to some quality (i.e., expertise). Beyond this, receiver, while retained, is considered more directly in the *Expertise Locator* system design discussed in Chapter 7. There, receiver (i.e., end user) serves to assess *Expertise Locator* retrieval output and may provide feedback (i.e., email notification) as part of *expert selection* (Chapter 2).
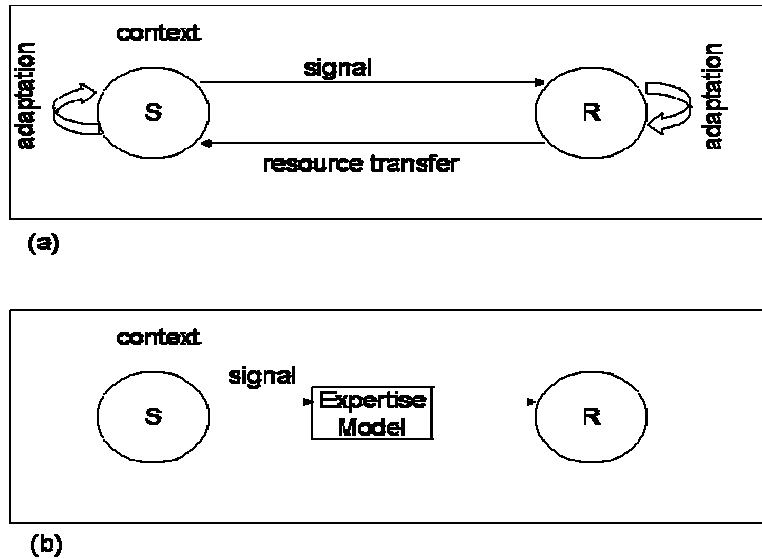
Figure 6-1: Simple Signaling Model (*S*, Signaler; *R*, Receiver)

Signaling is situated; so that signaling carried out within a particular organizational setting composed of specific activities, actors, rules, policies, and other elements that shape work. In that regard, signaler is embedded within an Activity Theory (AT) framework, as shown in Figure 6-2, and described in Chapter 4. Here, consistent with AT the subject refers to signaler and is a particular actor or group whose point of view is taken. Each activity has a motive (object); for example, to alert target groups to an actor's expertise in nanotechnology; that is the motive is to signal expertise. Instruments are internal or external mediating artifacts that transform the objective into an outcome. Artifacts are used to shape activity (e.g., use of a search engine to locate information) but they are also the product of activity; i.e., a retrieval list or formal report.

The community is made up of other actors who share the objective with the subject; this may be all ListServ members, a project team or meeting attendees. As such, artifacts and community context are effectively signaling evidence; here defined as artifact and social signaling evidence. Rules shape actions taken by the subject and interactions within the community. Some rules may be formally mandated others built up informally from group consensus. The division of labor determines how the activity is distributed across the community. For example, within a particular activity system, such as a project, there may be multiple signalers and in some cases multiple signalers associated with the same artifact; as in multi-authorship documents.

116

While activity theory provides a rich conceptual framework for designing collaborative spaces or analyzing extant work contexts, it is used here to inform the design of an Activity Space (AS), useful for expertise detection. In particular, the expertise model developed here incorporates activity spaces; sampling frames that capture signalers, signals, and organizational work context consistent with the reduced signaling game model introduced in this chapter. The selection of specific activity spaces for use in the expertise model is discussed in Chapter 5.
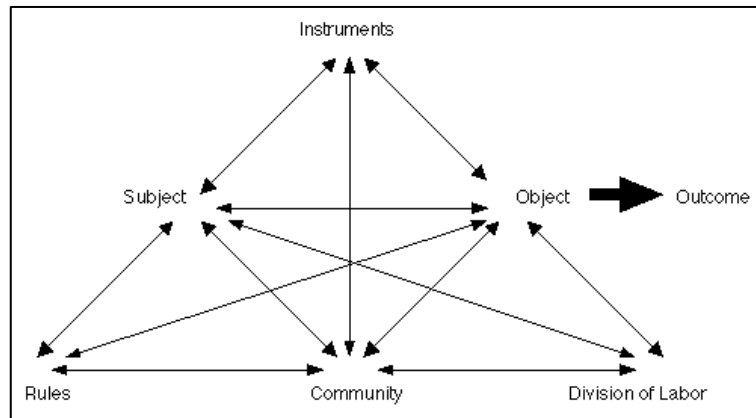
Figure 6-2: Activity Spaces and Expertise Network

6.2   Model Concepts

Within a domain, experts establish credentials, build reputation and trust through structural and relational embeddedness, for example, Granovetter, (1973). That is, they tend to work (and signal their expertise) within groups or communities-of-practice consistent with their area of specialization. As such, in the model developed here, an actor can be associated with specific work products (artifacts) and can be viewed as embedded within a particular activity space (e.g., a ListServ or a project) so that both artifacts and social interaction are viewed as signaling evidence and can be aggregated within a particular work setting to produce expertise ratings. The underlying concept is illustrated in Figure 6-2. The bipartite (2-mode) affiliation graph at the top shows membership across four activity spaces. Here activity spaces are noted by rectangles and actor nodes by circles. For example, actors A, B, and C are part of the same activity space context (1); however B and C are also members of a second context (2) and,

therefore, they overlap with D, E, and F. From this perspective actors may be characterized by their links to activity spaces[97]; that is, links to work context.

The 2-mode to 1-mode transformation at the figure bottom shows *actors linked to actors* based on activity space co-membership. The activity space context is represented here using shading. For example, activity space (1) and activity space (2) overlap since nodes B and C are members of both spaces. The key point here is that activity spaces provide a contextual overlay on the global graph structure imposing "locality" from which the importance of nodes in the network can be computed.

More over, the application of activity space weightings matches up with how users weight expertise evidence based on context. For example, a paper delivered to a prestigious conference may be weighted higher than one presented at a lunch time seminar; other factors not withstanding. It is important to note that assigning local activity space weightings to nodes does not preclude use of global weighting as well; and in that regard the model outlined here is quite flexible.

Bipartite Activity Space Graph

Single Mode Graph: Expert Linkages based on Activity Space Co-Membership
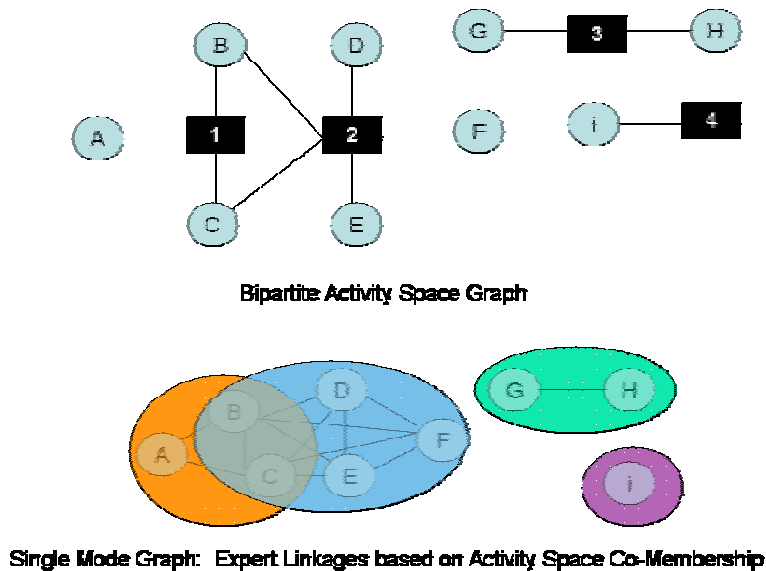
Figure 6-3: Activity Spaces and Expertise Network

---

[97] For simplification, actor attributes, artifacts, and events are not represented in this view but are integral to the overall model described below.

The single mode graph in the bottom of Figure 6-1, suggests ranking expertise based on nodal importance; where importance is based on graph structure and nodal attributes. White and Smyth (2003) discuss a general framework that subsumes various classes of ranking algorithms such as weighted paths and Markov Chains. The HITS algorithm, Kleinberg (1999), and PageRank, Brin and Page (1998), as well as a number of variations of these algorithms have also been used to rank vertices. In general, those algorithms exploit network structure in computing nodal importance. Jin and Dumais (2001) combine a content-based score with a link-based score to determine an overall node score with regard to a query. They use a spreading activation-like model over the link structure to compute a final network score. The algorithm proposed here uses artifact (signaling) evidence (where artifacts may be relevant documents, project charges, awards, etc.) as well as graph structure (social signaling) to determine overall ranking. However, the algorithm differs from Jin and Dumais in that local context derived from activity spaces influences nodal priority; in effect neighborhood evidence weighted globally is used to compute node importance. Referring to Figure 6-2, the importance of node "B", for example, is a function of artifact evidence, the connectivity between "B" and related actors, and the global significance of the contexts that "B" is embedded within.

The populated evidence space is viewed as a series of table pairs, Figure 6-2, in which each activity space has an artifact and social evidence table. The rows are subspaces within each activity space and the columns are actors; i.e., candidate "experts". Therefore, the first cell in, say, the artifact table, is the artifact weight for actor "1" in subspace "1" of activity space "1". Taking, say, the project activity space and the artifact table, this is the weighted score from 5 documents (artifacts) associated with actor "Stephen Sandina" in "Project Rome" (a specific project subspace). Below, this multi-table evidence aggregation scheme is formalized as part of the expertise ranking model. Each activity space essentially assigns scores used to rank actors and a fusion algorithm combines the separate rankings into a composite ranking.
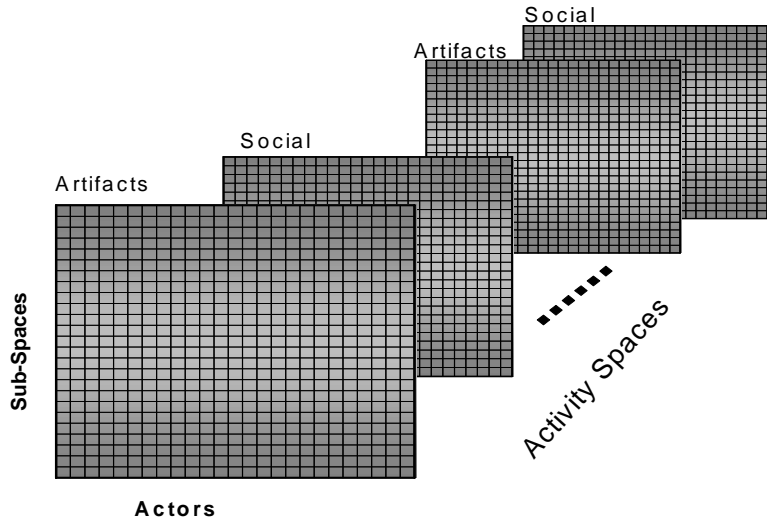
Figure 6-4:   Multi-evidentiary Tables

6.3   Mathematical Model

Architecturally, the system is viewed as a multi-agent decision model in which each activity space is associated with a decision agent, Figure 6-3. Decision agents support evidence collection, synthesis, and actor ranking specific to an activity space. For example, the ListServ decision agent retrieves message posts related to a query, parses headers, extracts routing metadata, and, potentially, addresses more complex processing like handling discussion threads or extracting named entities (e.g., locations and personal names). Here artifact evidence (postings) are accumulated in one evidence table, and social evidence (i.e., who communicates with whom) is stored in the other table (as illustrated in Figure 6-2). The ListServ decision agent weights and aggregates both kinds of evidence across all relevant actors.
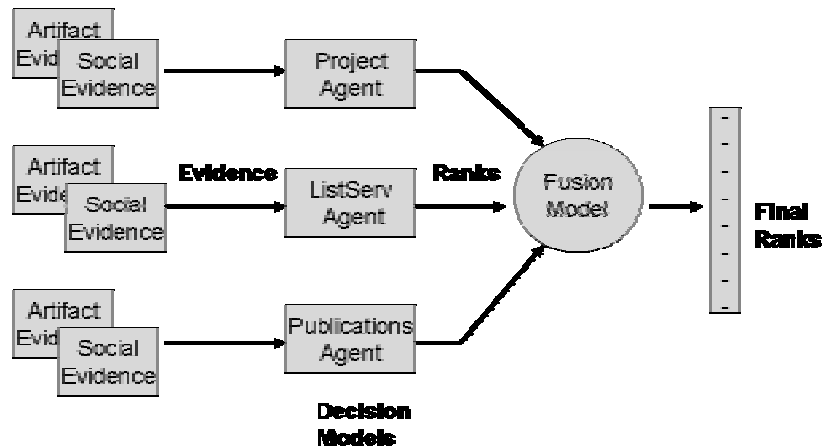
Figure 6-5: Fusion Model

Actors are scored and ranked based on their cumulative weighted evidence. A similar process is used by each decision agent; however, individual agents work on independent sources, and are tailored to address the characteristics of a specific activity space. The combination of expertise rankings and co-work relationships derived from activity space membership suggests viewing the overall retrieval as a type of expertise network or graph in which actor ratings are equivalent to *nodal importance*. This is depicted in an actual retrieval output, Figure 6-4, where the node size reflects expertise score (i.e., nodal importance). Co-work (query-relevant co-membership in ListServs, formal organizations, and projects) is reflected in the edge connections between nodes.
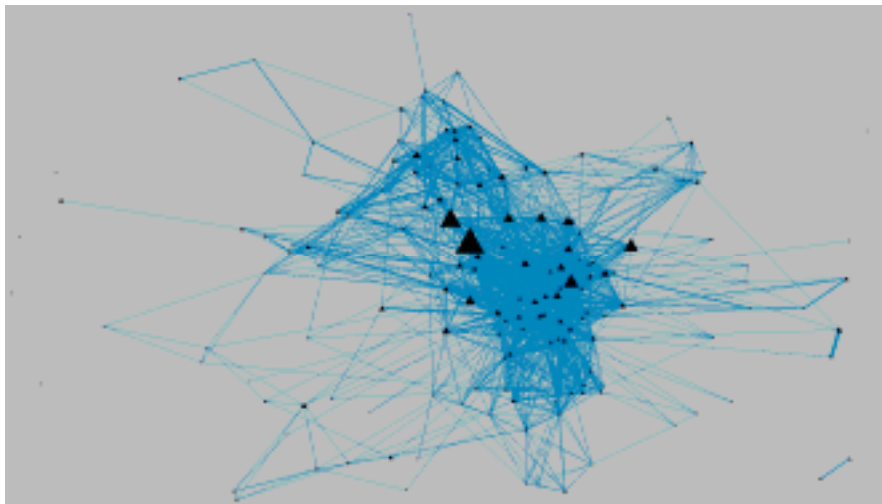
Figure 6-6: Expertise Network

In general, nodal importance (expert score) is computed as:

$$I(p \mid q) = \sum_{\forall i} \alpha_i \cdot E_{i,\bullet,\bullet,p} \qquad (6\text{-}1)$$

where $I(p \mid q)$ is the importance of person, p, for query, q; α_i is the weight assigned to activity space, i; and $E_{i,\bullet,\bullet,p}$ is the aggregate (artifact and social) evidence for all subspaces[98] within activity space, i. For a particular activity space, i, and person, p:

$$E_{i,\bullet,\bullet,p} = \sum_{\forall k} (\omega_{ik} \bullet \sum_{\forall j} \beta_{ij} \bullet e_{i,j,k,p}) \qquad (6\text{-}2)$$

Here, $e_{i,j,k,p}$ is the evidence of type k, found in activity space i and subspace j, that is associated with person, p. Then, β_{ij} is the weight of importance assigned to subspace, j in activity space, i. There are k signal evidence types so that the total subspace evidence for each evidence type is scaled by ω_{ik}, the weight assigned to evidence type k[99]. The weights provide a basis for biasing the importance ratings for one type of signaling evidence over another (ω_{ik}), and for treating some subspaces as more important than others (β_{ij}). Activity space weighting, α_i, is used in the fusion algorithm discussed in following sections.

Basic model components are presented visually in Figure 6-5. As shown, the expertise rating for person, p, and query, q, is the weighted aggregate evidence from each activity space. As discussed the model supports fine-grained weighting (used optionally) to assign varying weights as to activity spaces, evidence types, and individual subspaces within. This allows, say, document evidence to be weighted higher/lower than social evidence for selected subspaces. In addition, subspaces may be weighted according to their discrimination value so that reports generated by internal research projects may be given higher weight than reports generated by business planning tasks.

---

[98] For example, a subspace might be a particular project within the project activity space.

[99] There are two evidence types in the current model: artifact and social. Therefore, k=1, 2.
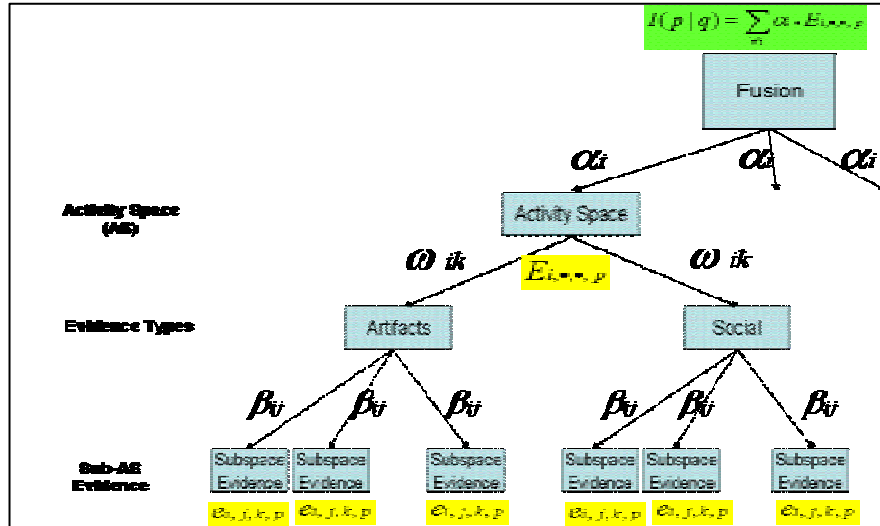
Figure 6-7: Basic Model Components

The model as represented in Equation 6-1 is simply the weighted aggregate evidence across decision agents; where each agent is a ranking function operating on an activity space. To provide a normalized basis for combining evidence, agent scores are converted to ranks. For example, the ListServ agent outputs a ranked list of actors based on the evidence associated with the ListServ activity space. This carries through for each agent. Rank distribution and ranking transformations have been addressed in research related to collection fusion for example, French et al (1999), Fox and Shaw (1994), Bartel et al (1994), and voting schemes Lifantsev (2000), Montague and Aslam (2002). Here we incorporate the CombMNZ weighting scheme Fox et al (1993) into the overall weighting function. CombMNZ adjusts the score to account for the number of activity spaces that each person has evidence in, and Borda[100] counts are used to transform retrieval ranks. Therefore, the nodal importance score from Equation 6-1 is transformed into a ranking function as follows:

$$I(p \mid q) \Rightarrow I_R(p \mid q) \qquad (6\text{-}3)$$

where

$$I_R(p \mid q) = N^\gamma \bullet \sum \alpha_i \bullet B(E_{i,\bullet,\bullet,p}) \qquad (6\text{-}4)$$

---

[100] In the Borda scheme actor ranks are based on the cumulative points across voters. The top ranked actor from each voter is given c points, the next c-1, etc. Unranked actors are given points based on an equal distribution of remaining points. Interesting properties of Borda Counts are described in Saari (1999).

Note that α_i is from Equation 6-1, above, $B(E_{i,\bullet,\bullet,p})$ is the Borda count computed on activity space, i, and N^γ is the number of populated activity spaces to the power (-1, 0, or 1), as in CombMNZ. If the power is *-1*, then a simple average is computed. If the power is *0*, the activity space count is not used, and if it is *1*, the sum is scaled by the number of populated activity spaces.

Note the model as specified in Equation 6-4 is similar to the Weighted Borda-fuse model, Aslam and Montague (2001), although the underlying activity-space based evidence aggregation is qualitatively different. Another key distinction is that the fusion method used here requires no training as discussed in the next section. CombMNZ, used optionally, adjusts final ranks to account for evidence distribution. Next, rankings are modified to reflect organizational attributes associated with a particular candidate.

The model as specified in Equation 6-4 does not explicitly account for actor status outside that attributed to signaling evidence associated with a particular query. This suggests that when two actors have roughly the same signaling evidence there is no distinction between an actor with lower enterprise status and one with higher prestige. Interestingly, a number of users suggested that information regarding an actor's organizational position, role, tenure, or affiliations should be considered in any final ranking. With that, the model was extended to incorporate a role-based weighting for each actor independent of the query; the implementation is described in Chapter 7.3.6.

R_p is the role-based status scaling for person, p, and the expertise ranking model is

$$I_R(p \mid q) = R_p \bullet N^\gamma \bullet \sum \alpha_i \bullet B(E_{i,\bullet,\bullet,p}) \qquad (6\text{-}5)$$

As such, an actor's overall expertise rating is a function of total evidence, $\sum \alpha_i \bullet B(E_{i,\bullet,\bullet,p})$, with optional weighting based on evidence distribution, N^γ, and actor's role status, R_p. Essentially, given specific evidence, experts with a larger organizational "footprint" and higher corporate recognition are rated higher; all other factors constant. Model outputs can be viewed as a 3-

dimensional landscape reflecting the distribution of signaling evidence (scores) across activity (social) spaces and actors. This is illustrated in Figure 6-7, where evidence associated with (N=65) actors is distributed across the top 100 subspaces within a particular activity space. The next chapter discusses Expertise Locator Prototype implementation.
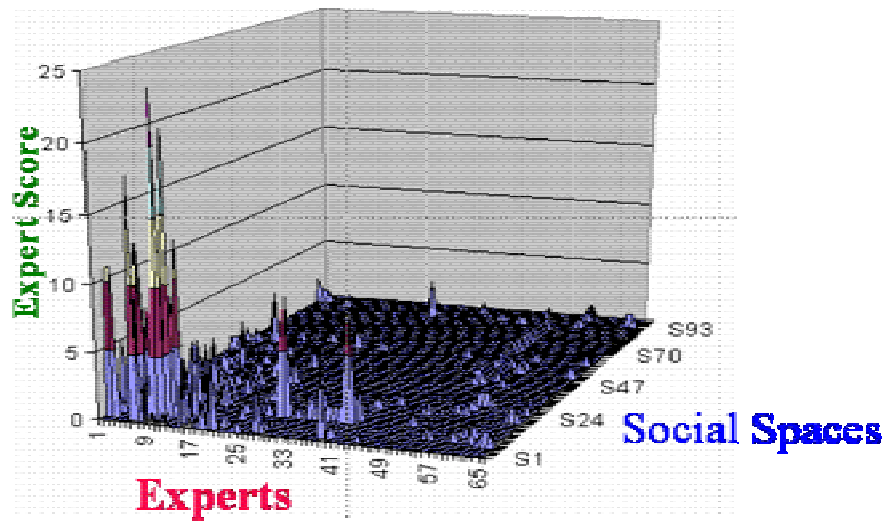
Figure 6-8: Expert Scores Distributed across Social Spaces

7 *Expert Locator* Prototype

This chapter describes the *Expert Locator* system architecture, user interface, and functionality. Specific emphasis is given to systems engineering issues and design tradeoffs central to deploying the prototype into an operational environment while still maintaining design flexibility needed to support this research.

7.1   General Architecture

Expert Locator is built around an extensible expertise model used to combine evidence from multiple sources and work settings. Conceptually, the system operates as a distributed information retrieval system, supporting evidence retrieval from disparate enterprise collections and services. *Expert Locator,* developed in the Perl and Java programming languages, utilizes a backend Microsoft SQL Server database populated with various data including project, ListServ, and directory information obtained from the corporate LDAP server. This database complements other information that is dynamically retrieved from corporate search engines (e.g., Google) in support of a particular query. The prototype is designed to not duplicate information in the *Expert Locator* database that is otherwise readily available from existing corporate search engines. The system architecture, Figure 7-1, consists of four major components: **User Interface**, **Evidence Collection, Expertise Model**, and **Results Output**.   Each component is described below.

7.1.1   User Interface

The user interface supports simple and complex user queries. Queries may be entered as free-text, or through a multi-field form allowing for multiple parameter settings and special filters used to adjust retrieval operations. For example, users may restrict retrieved experts to those from a particular geographic location or organizational unit. Users may also control search depth; restricting the search space to selective activity spaces, or constraining search engines to some maximum number of retrieved items.   Some of the advanced search options were initially

used to support research investigations and carryover from a special research version of *Expert Locator* that provided debugging support, metrics, and evidence characterizations.
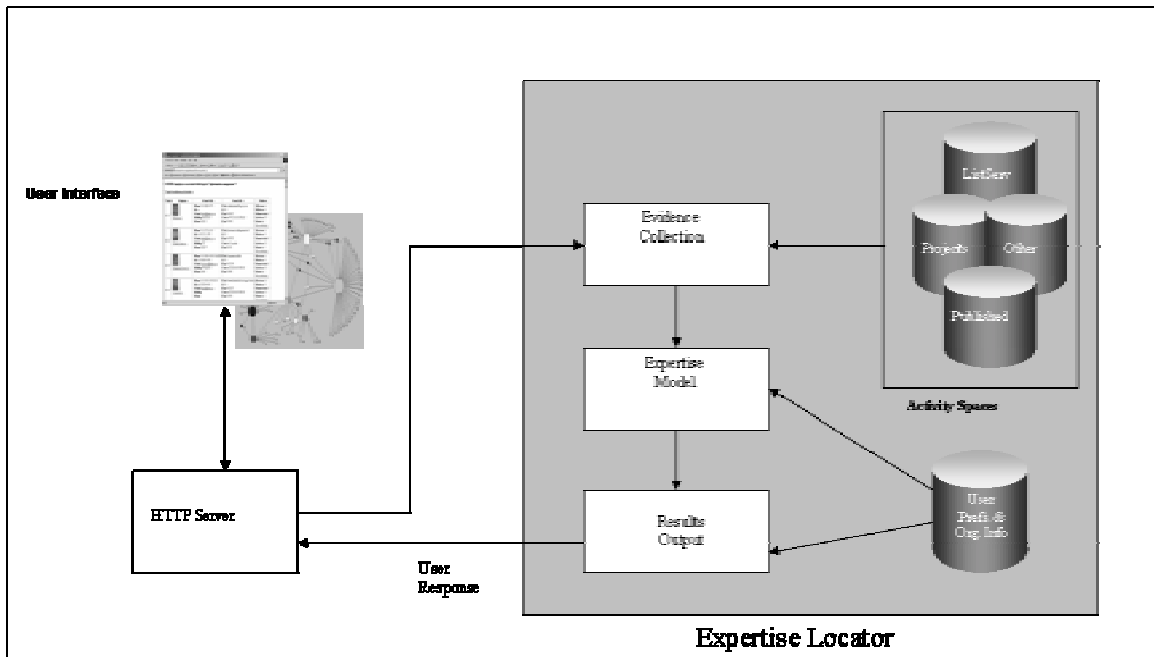
Figure 7-1: *Expert Locator* Architecture

As shown in Figure 7-2, *Expert Locator* supports simple query entry box (left), and an advanced query interface (right). Users tend to use simple keyword queries, but often use phrases or simple Boolean operators to increase specificity. Here query entry is made to mimic Google, the enterprise search engine; *Expert Locator* syntax is identical with that supported by Google. For example, users can use several keywords or form phrase searches such as: "*natural language processing*" AND "*data mining*".

The advanced query interface supports forms-based searches providing users direct control over several search parameters to augment standard Google-like query expressions. A short description of the main user-controlled search parameters is shown in Table 7-1. In particular, the parameter/option space may be divided into organizational filters used to constrain searches based on actor's organizational work context, and various settings used to control evidence weightings.
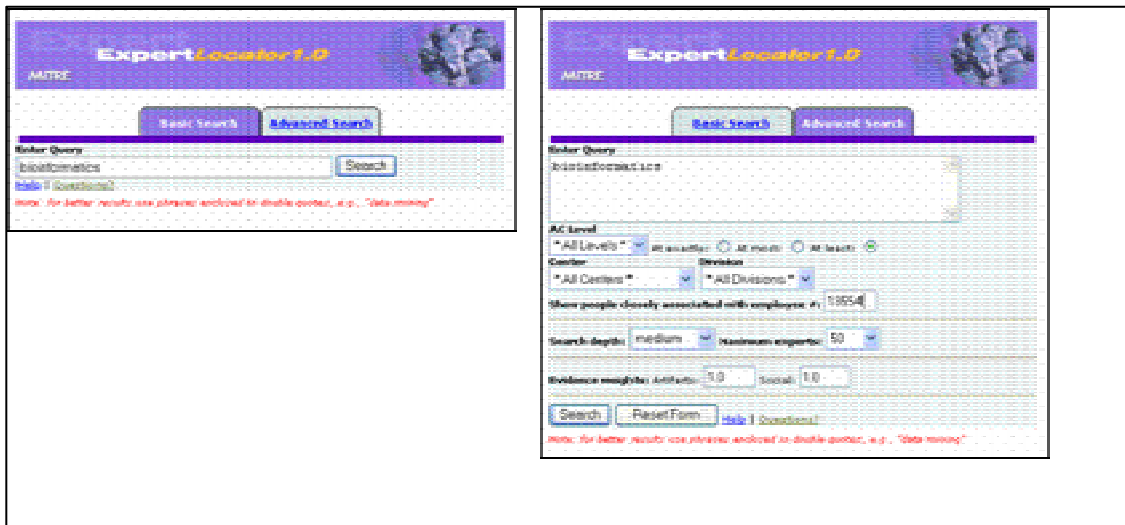
Figure 7-2: Basic and Advanced User Interfaces

As shown in Table 7-1, users can restrict retrievals to those from certain organizational units (i.e., Center) or specific AC Level (this is similar to professional status). In addition, users can control Search Depth (essentially a retrieval cutoff) in terms of the numbers of documents retrieved; this affects search time and possibly precision or recall.

Advanced Query Option	Operation Performed
AC Level	Retrieved experts are restricted to those having specific AC level with 3 ways to set the threshold: exactly the levels specified, at least, or at most the level specified.
Center	Retrieved experts are restricted to the work Centers selected
Division	Retrieved experts are restricted to the work Divisions selected
Search Depth	Sets the maximum number of retrieved documents from a particular search engine
Maximum Experts	Sets the maximum number of retrieved experts after fusing individual activity space rankings
Evidence weights: Artifacts	The weight assigned to artifact (e.g., document) evidence
Evidence weights: Social	The weight assigned to social evidence (e.g., activity space density)
Show People Associated With	Retrieved experts are restricted to those organizationally linked to the specified person (users enter employee ID)

Table 7-1: Advanced Query Options

Final ordination is based on separate rankings from each activity space model. A fusion algorithm merges individual rankings into a final ranked list. While each space may contribute a relatively large number of candidates, the user can restrict the final ranking to a maximum number, Maximum Experts. In addition, users can adjust Evidence Weights individually,

128

affecting the relative importance of artifact and social evidence. This has certain advantages with iterative searching where users may glean that one type of evidence is more useful than another. For example, social evidence might be more valuable in finding experts heavily embedded in dense expertise networks with less weight on those working in isolation.

Finally, users can anchor the retrieval results around a particular person by using the <u>Show People Associated With:</u> option. Here users can restrict retrieved experts to those tied to a particular person and this may support user navigation to selected experts or relevant intermediaries. This may be operationalized as "show me the highest ranked experts who have work relationships or are organizationally linked to person X." This results in an egocentric network where the ego (or target actor, X) is viewed in relation to X's alters. There are two possibilities here; users may choose person X to be someone with known expertise in the target query or to pick X from the general enterprise. In the former case, the retrieval graph may be viewed as a type of personal network conditioned on alters matching the query. In the latter, it might show how specific experts link to someone who is not an expert. For example, the latter case may be viewed as: "which experts are linked to X; where X is known to not be an expert." Other approaches to this type of association-based navigation have been explored; for example, systems such as SocialPathFinder, Ogata et al (1999) and ReferralWeb, Kautz and Selman (1998) use name co-occurrence extracted from Web pages, organizational charts, and other artifacts as the basis for identifying associations that can be used to guide navigation to experts or intermediaries providing referrals.

Collectively these parameters address two aspects of expert finding; retrieval and selection. While retrieval performance is central to overall performance, users may benefit significantly from having assistance in selection. Selection tools can provide a way of reducing retrieval noise, and allow users to exploit organizational knowledge in contacting experts. In particular, anchoring retrieval on a particular person using the <u>Show People Associated With:</u> option directly supports selection; providing insight as to how a user may select a particular expert or, alternatively, use intermediaries to obtain help or facilitate access to someone. Based on informal user feedback, this may be particularly valuable to, say, junior staff members who may

be reluctant to contact a senior scientist or manager; however, they may wish to select "peers" linked to the target actor.

7.1.2   Results Output

The Output subsystem produces results in different formats, such as HTML, XML, or a text delimited file, enabling the system to be invoked through an interactive client interface or alternatively from a script executed as a background task, such as an application that feeds a database.   In addition, the system has as a Java interface to enable use of more advanced graphical capabilities providing better support for exploiting social networking capabilities. The Java interface utilizes the InXight Star Tree SDK[101], a graphical utility that provides a flexible interface for visualizing and manipulating networks. Star Tree supports a Java-based API that enables Web-based (e.g., Applet) and stand alone graphical applications.

A simple retrieval graph, Figure 7-3, shows the list of top ranked experts that meet the advanced search criteria and query topic—a more complete retrieval graph is discussed, below. Retrieved experts are ordered based on rank (1 to "n"), and color-coded from red to blue (hot to cold) to identify how similar they are to the query. The query is shown at the center (usually truncated for display purposes). There are several options (shown at the top of interface) that allow users to chose a particular display type, such as tree view, shown below. Other views are more appropriate for more complex graphs where users can control the overall layout more effectively and "hide" or display edges by simply mousing over the graph. In addition, users can color nodes based on various attributes such as home division, AC level, geographic location, etc.

---

[101] www.**inxight**.com, Accessed on December 20 2005
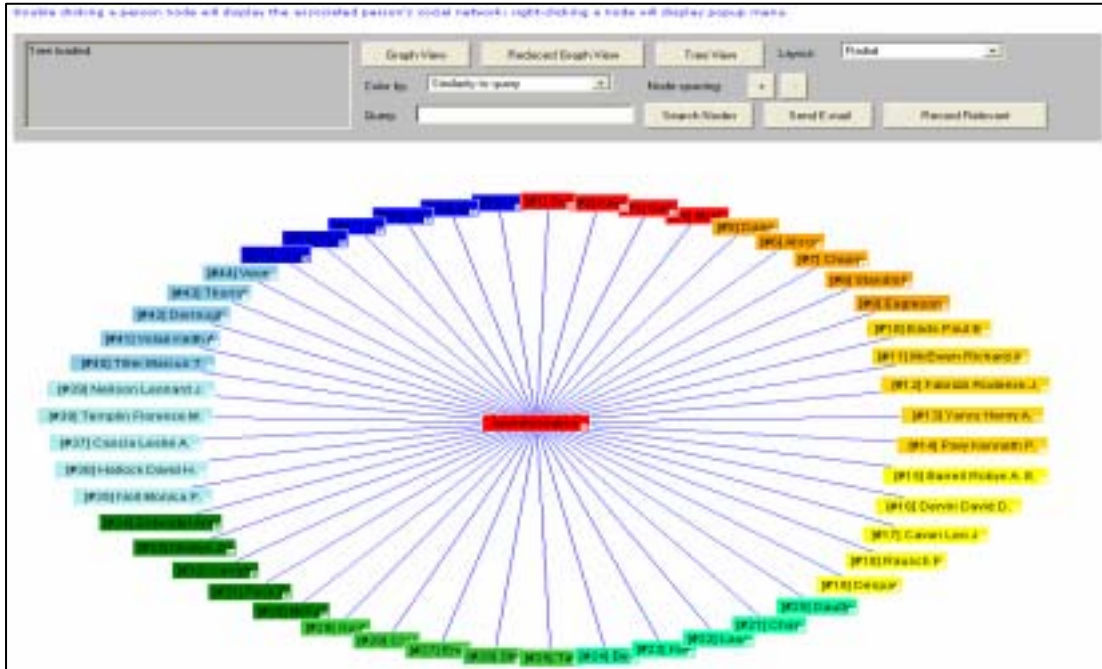
Figure 7-3: (Optional) Initial Retrieval Screen

The complete retrieval graph (tree), Figure 7-4, includes all actors and their work ties. In particular, each actor is linked to their associated activity spaces (represented as black rectangles), and in this view[102], a single actor may be assigned to multiple actor nodes; one node for each activity space he/she is linked to. Using the expertise ranking attribute for colorization, this view provides a quick way to find the highest ranked experts (warm colors) and their distribution across activities. This view can be easily changed to reflect the distribution of experts as to geographic location, home organization, or other attributes through nodal color coding. For example, coloring nodes by geographic location may reveal that top ranked experts are co-located at some remote site as opposed to corporate headquarters.

---

[102] Other display modes represent each actor <u>once</u> with multiple edges used to reflect membership in more than one activity. Other modes alternatively mask or display edges based on user interaction.

131
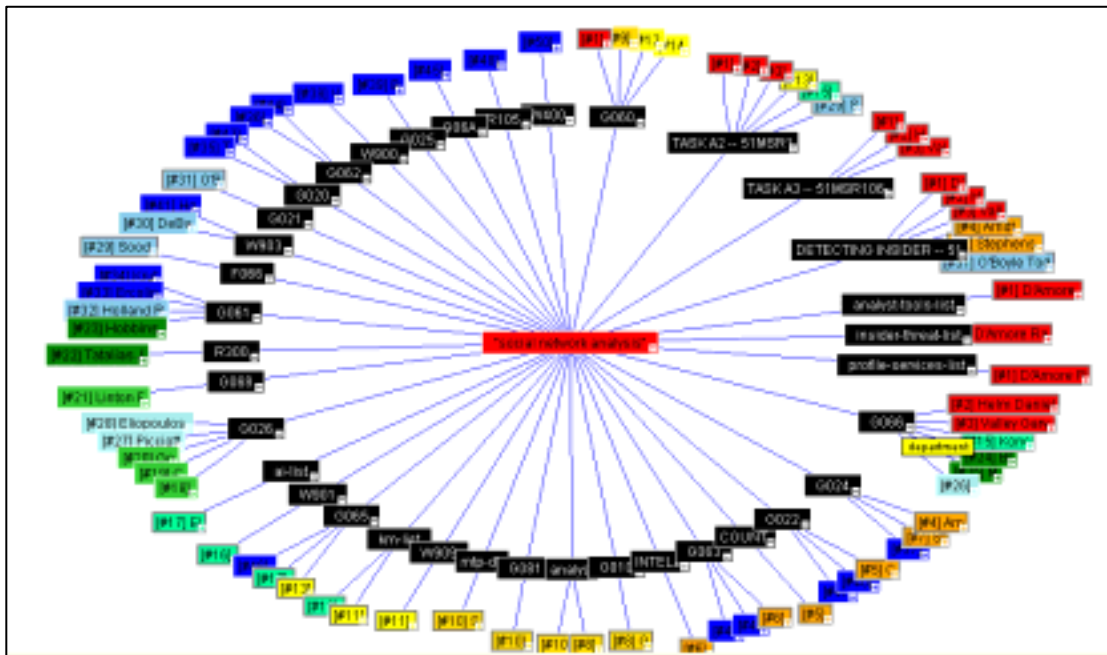
Figure 7-4 Expertise Network

Users can mark up the retrieval graph as part of a limited workflow capability. Two aspects of user interaction are selection and communication. While browsing the expertise graph users may tag certain experts for follow-up contact or to simply make a list for future use—possibly to support a meeting. A simple example, Figure 7-5, illustrates actor tagging for inclusion in an email distribution list. Note, the email icon attached to certain actors. Then users can send an e-mail to the tagged list from the *Expert Finder* interface without leaving the system. This has utility where a user wishes to send a particular request for help or send feedback to an individual or group in lieu of phone or face-to-face contact.

The system was designed to support collaborative searching, team generation, and relevance feedback. For example, planned extensions to the system will allow multiple users to collaborate on building a project team. In this use, each user generates queries representing multiple expertise areas as the basis for building a heterogeneous team. As part of post-retrieval analysis, users mark up retrievals as a form of nomination process. A backend database application merges nominations using various voting schemes. Team membership can be biased according to queries matched and various actor attributes such as location, professional status, or

132

organization home. Note that in Figure 7-5, selected actors are tagged as indicated by the check
mark, ☑.
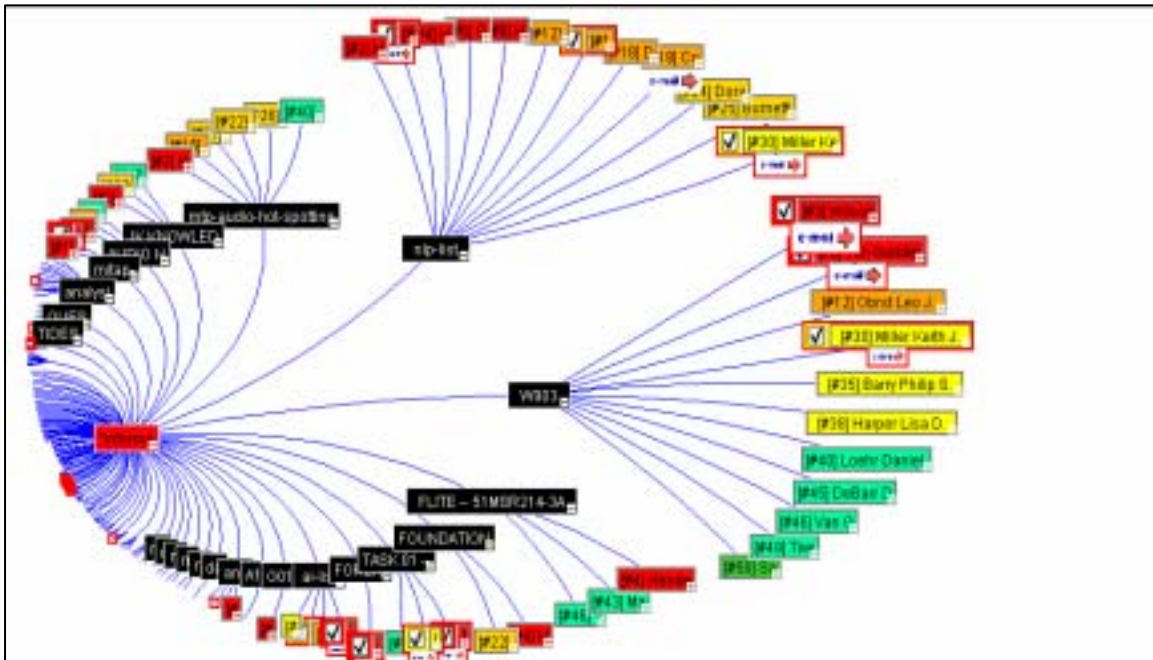
Figure 7-5: Retrieval Network with Markups

Actor tagging can support a range of other applications. For example, tagging may be used as
part of relevance feedback; where significant actors are used in a query-by-example mode to
adjust query terms or retrieval parameters. In one case, user feedback may be used to adjust the
relative weight assigned to evidence types or to modify the importance of one activity space over
another. More generally, tagged actors linked to historical queries can be used to generate an
expertise directory and recommender system. The system could build a query history, record
actor evaluations, and support expertise queries against a sort of dynamic directory. The
directory could be used as a complement to *Expert Locator* or as a separate browsing service. A
flexible relevance feedback option based on user markups of artifacts or actors is a planned
future study area.

Actor metadata may be obtained by right-clicking on a particular actor's node. Various types of
metadata may be served up to include personal information found in the corporate LDAP server.
In addition, an actor's artifact evidence may be displayed; this may include various document

types to include publications, ListServ posts, and items from share folders as represented in Figure 7-6. This is one source of evidence used to generate expertise ratings. Users can easily download these items for inspection.
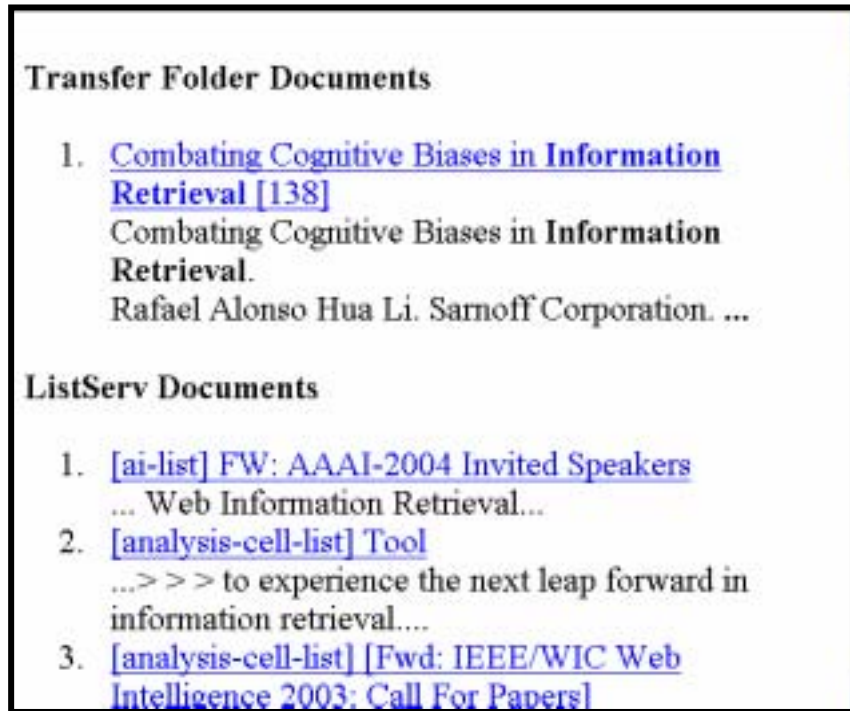
Figure 7-6: Document Artifacts Associated with an Expert

While most interaction has been at the level of the whole retrieval graph; users can also drill down into personal networks. Double clicking an actor node will display an actor's personal (ego) network in a new window, Figure 7-7. A personal network contains an actor's nearest neighbors; other actors relevant to the query that have co-work relationships with ego. From a social network perspective, a personal network may be rendered as a bipartite graph that shows actors linked to activities. Here, ego is linked to associated activity spaces along with co-members. This is similar to authorship graphs where activity spaces are equated with authored papers, ego is an author of interest, and alters are co-authors with ego. The system provides options to manipulate the social network graph similar to those found on full retrieval graph. Using the tree view, all actor links to activity spaces are visualized. Using the graph or reduced graph modes each user node is represented only once, and multiple edges are used to show a

user's ties to several activities. These options provide a way to declutter the graph, especially when there are large numbers of actors and activity spaces and where there are many instances where actors have multiple activity space membership.

The default nodal coloring shown on a personal network reveals how similar altars are to ego. Here a simple social correlation measure computes "social distance" as, 1-d, where "d" is the fraction of activity overlap between ego and alter. Warmer colors (e.g., red, green) identify actors that have highly similar work patterns; while cooler colors (e.g., blue) identify actors with weaker work ties to the central actor. Color encoding, as used here, provides insight into additional experts that can be contacted along with (or instead of) the central expert.
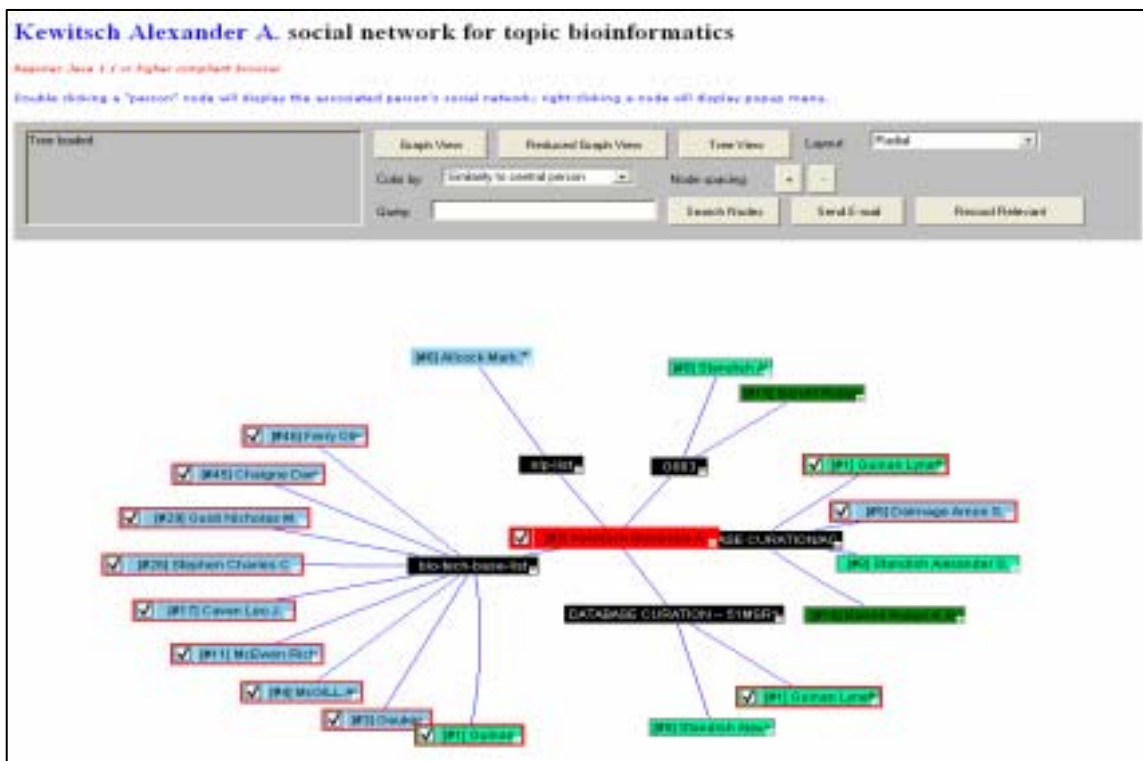
Figure 7-7: Personal Network

Overall, personal networks generate (local) social context for each expert; identifying activity spaces and other actors that a target actor is directly associated with. A user may use personal networks to identify how tightly connected a particular expert is to others, or to determine which actors might serve as intermediaries to, or as a surrogate for the target expert.

7.2   System Processing

This section addresses system processing; to include high-level operations such as evidence retrieval, and lower-level methods and sources used to instantiate the expertise model. Implementation issues arising from integrating the system in an operational environment are addressed throughout this section; especially with regard to certain tailoring needed to accommodate available resources, use cases, or policy.

7.2.1   Evidence Retrieval

While flexible query, results visualization, and user interaction are keys to effective usage, evidence retrieval and synthesis are central to overall retrieval effectiveness. The starting point here is the enterprise. The MII corporate Intranet is a heterogeneous environment made up of disparate collections and information services. In most cases the services are managed as part of the corporate infrastructure and that precludes re-hosting core capabilities to support new applications, and discourages heavy usage of operational systems that may degrade the quality of service provided to general users. As such, the collection and access strategy used to support *Expert Locator* involves tradeoffs designed to minimized impact on enterprise operations; this includes periodic project data collection and ListServ real-time capture.

Evidence collection is viewed as a distributed retrieval operation where disparate enterprise services and collections are accessed using expertise queries. The system is distributed in that key artifacts or social relations may be embedded within multiple autonomous systems such as a project database, meeting and calendaring services, and the corporate search engine.

A search broker manages the distributed search operations[103], Figure 7-8. The Google enterprise search service supports searches against "formal" publications (e.g., white papers, project reports), and publicly shared files (e.g., briefing slides). Formal publications are submitted by

---

[103] This represents system architecture during the bulk of testing; later corporate search services were extended so that Google provided access to all searchable artifacts. Using Google, partitions are used to segment retrieved documents into activity spaces so that evidence can be correctly counted in the Expertise Model.

users using a procedure that ensures specific metadata are associated with each item; for example, author. The Google web crawler collects these items based on a crawl schedule specified by corporate system administrators and not driven by *Expert Locator* requirements. Share folders consist of a wide range of "informal" products; however, documents that are deposited into a user attributed share space without a supporting metadata extraction process may not be easily assigned to specific authors.

The query is also directed at ListServ posts stored in the *Expert Locator* backend database. As a result, items retrieved from Google, and *Expert Locator* database are used as artifact evidence by the ranking model. The query also retrieves attribute data on each person. However, relational evidence, to include actor-activity or actor-actor linkages is generated by the model from analysis of simple artifacts and activity space data.
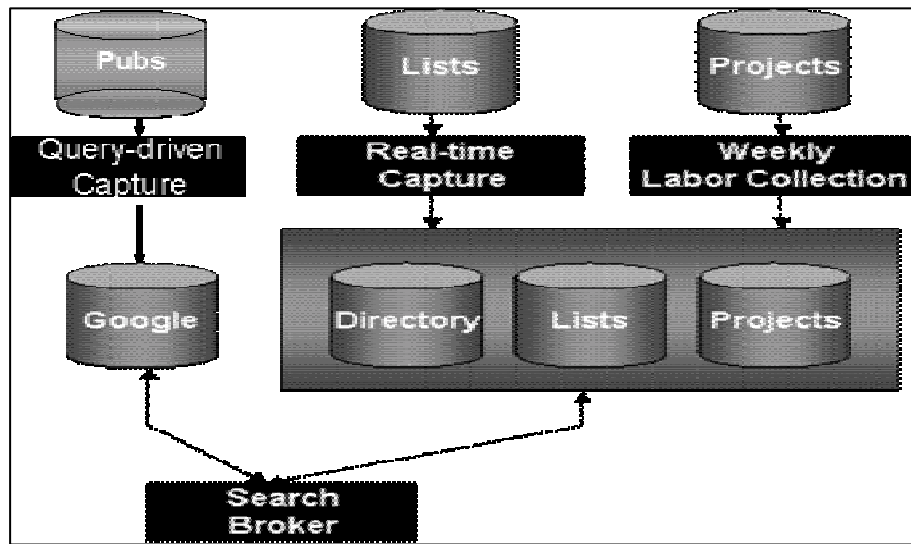
Figure 7-8: Hybrid Data Collection and Access Architecture

There are two back-end data collection operations used to feed the ListServ and Project data store maintained by *Expert Locator*. A relational database is setup to warehouse postings from each ListServ. The database includes header information from each posting (such as TO, FROM, and DATE) and the full text of each post. There are several thousand ListServs archived for public use on the MII. Each ListServ typically has scores of users and some have more that

137

500 members. In order to obtain ListServ postings in near-real time *Expert Locator* essentially subscribes itself to each archived ListServ; this is done with permission from ListServ "owners". Collected postings from each ListServ are then used to update the ListServ database[104].

Project data are collected on a weekly basis.   Metadata for each project (such as *Project Title*, *Task Leader*, and *Project Number*) are collected from HTML pages. Labor charged by each project member is obtained and appended to historical data to keep a rolling count of total hours charged. This provides a basis for tracking resource utilization across members as well as over time. Time spent working a project may be used to filter people from the project team or it may be used to weight the importance of each team member. Finally, most of the directory services data are also stored in the database for use by the detection algorithms. Staff photos and other data used on output are dynamically retrieved during query execution.

7.2.2   Collection Processing

Evidence retrieval is a hybrid process involving periodic collection, as well as dynamic query-based access to multiple enterprise systems.   In general, various actor or activity space attributes are collected on a periodic basis or may be event-driven in response to organizational events such as a structural change in the organization. Viewed as a series of "snapshots", this approach scales reasonably well when various attributes are relatively static or slowly changing; for example, an actor's home department, project membership, and geographic location will tend to vary little across weeks or months. This background evidence is combined with query-specific evidence in which relevant artifacts and social context are "collected" as part of a retrieval process. Both collection modes must by synchronized in order to support an *Expertise Locator* query but equally important, the processes must work within an operational environment that imposes a number of access and resource utilization constraints on *Expertise Locator* operation.

For example, project data collection is handled periodically as a batch update (independent of queries) and is scheduled so as to reduce impact on MII performance. *Expert Locator* collects

---

[104] The Google search engine did not index ListServ postings during the period this research was conducted. As such, separate indexing, storage, and retrieval had to be provided to incorporate postings into the overall retrieval.

project membership lists and labor charges; potentially filtering out actors that have negligible involvement on a particular project. The system collects weekly project labor hours from each project task; on the order of two to three thousand tasks per year. However, it is costly to perform weekly updates as it requires running a collection script that accesses project data across all MITRE contract bases, extracts labor charges[105] per person per task, and then updates several database tables. This is exacerbated further by constraints imposed on access to relevant corporate databases. Full access to corporate data is restricted to users satisfying need-to-know constraints. Therefore even though labor data may be viewable on a staff member's corporate web page the same data may not be accessed through the corporate database since that increases the risk of access to restricted fields like *salary* or *date-of-birth*. This necessitated a more lengthy process in which project labor data are extracted from the publicly viewable Web page, a process known as *screen scraping*[106]. While this circumvents privacy or need-to-know concerns it degrades the collection update process since screen scrapping incurs significant file access costs.

To further reduce impact on MII resources it was necessary to explore update cycles that were less frequent and in particular to assess the impact of longer update intervals on *Expert Locator* performance. System testing suggested that the *Expert Locator* was fairly robust to the update interval; in most cases, changes in labor hours from week to week did not significantly impact *Expert Locator* search results. If the update is run bi-weekly or even monthly there is very little degradation in expert rankings for a given query. To better ensure that *Expert Locator* could adapt to corporate policy changes, a hybrid update scheme was developed that supports weekly updates when feasible, shifts to longer update intervals when mandated, and inserts event-driven updates to ensure that significant work perturbations are reflected; say when new business models affect project labor distributions or when internal research projects are awarded.

---

[105] Labor is recorded as "hours worked"; not as salary expenditures.

[106] Screen scrapping is a text extraction process that strips out relevant text segments from HTML pages.

7.2.3   Resource Re-hosting

Privacy, intellectual property, and MII resource utilization were key factors that shaped the *Expert Locator* information collection and access architecture.   For example, corporate policy precludes storing published documents (already managed by an enterprise search engine) for re-indexing by a dedicated *Expert Locator* search engine. This has both strategic implications regarding system use as well as tactical issues regarding system performance. For example, use of the MII (Google) search engine precludes efficient integration of named entity extraction into *Expert Locator* analysis. Entity extraction implemented as a post-retrieval operation is inefficient while integration into the MII search engine is prohibitive since this would require modification of Google's low-level indexing operations and data structures. The net effect is that tight reliance on the MII search engine precluded developing certain pre- or post-retrieval strategies; however, it did facilitate rapid integration of *Expert Locator* into the MII Intranet environment.

7.2.4   System Responsiveness and Design Choices

Response time for a given query is determined by network overhead, query specificity, search depth, model computation, and results presentation. Some factors, such as network loading, are largely outside the control of the system since they depend on enterprise network traffic and loads on specific servers. In other cases, performance is dictated more by design and intended use. For example, the system downloads a complete retrieval graph; this is costly on first instance, but subsequent analysis and browsing can be done quite rapidly with little latency. As such, browsing, evidence perusal, and visualization benefit from local caching but users pay a front-end cost to retrieve data and generate needed back-end context. However, this tradeoff aligns with both the research and operational direction planned for in that the perceived value of the system is in back-end, post-retrieval operations. Users will have the ability to rapidly browse retrieval results, evaluate supporting evidence, and explore personal networks with little latency.

Search efficiency and accuracy are complex performance issues, especially when viewed from the perspective of user directed searching. Here, user control over searching has significant implications as to overall retrieval time, results composition, and precision/recall. A particular search setup may be consistent with user's internal view of what constitutes relevance; however, this may not be "optimal" when juxtaposed with some other performance measure. Said otherwise, users may use the system in ways that are not optimal. This presents a quandary in terms of how much control to give users in terms of, say, restricting the search space or limiting the types of evidence used.

Search constraints were viewed along several dimensions: coverage, completeness, and evidentiary types. Here coverage relates to search breadth and completeness is associated with search depth. Each of these search aspects can be operationalized and given system-specific definitions:

- **Coverage**: users can restrict the search space by reducing the number of work contexts examined by the expertise model. Here users can select the number of activity spaces to search, analogous to the number of collections targeted for retrieval. This can have significant affects on performance since a user may decide to use, say, only the ListServ activity space which would likely affect recall; restricting the retrieval to some subset of all relevant actors. On the other hand, selecting, or eliminating, specific collections can bias the system to certain behaviors and artifact types which can be useful when searching for certain niche expertise or restricting evidence to certain "transactions". However, as noted, this can lead to suboptimal performance when user intuition does not match where and by whom relevant work is actually being performed.
- **Completeness**: Users can adjust search depth by setting a retrieval cutoff. The cutoff can be set to *Low*, *Medium*, or *High* system-fixed values, and applied so as to limit the number of items searched in each activity space. Changing search depth has separate affects. With regard to artifacts, relaxing the cutoff from, say, *Medium* to *High* will generally increase recall but reduce precision. That is, retrieving more items may increase retrieval "noise". However, relative to expert rankings, the retrieval cutoff may

not necessarily affect the final expert ranking; that is, final expert ranking may be cutoff-independent[107].

- **Evidence weightings**: Here, users can shift emphasis between artifact and social evidence. From an expertise signaling perspective, artifact and social evidence are viewed as signal types or expertise "advertisements". Therefore, users may wish to assign greater importance to, say, artifact rather than social evidence.

Overall, the system was engineered to include selected search controls so that users could directly adjust search behavior; either through restrictions in evidence types or work context (activity spaces). In other cases, search restrictions or performance adaptation was built-in as a system option. Most of this is motivated by the need to support the user or system administrator over long-term use. However, the affect of restricting search to various combinations of activity spaces (coverage) was directly addressed in the evaluation carried out in Chapters 8 and 9.

7.3 Instantiating the Expertise Model

This section focuses on evidence sources and processing used to instantiate the expertise model covered in Chapter 6. Evidence sources (e.g., ListServ postings) retrieved from a user query are transformed into expertise model inputs. Essentially, sources are transformed into evidence records (ER) which are mapped to artifact and social evidence tables processed by the expertise model. This discussion is intended to provide sufficient guidance for researchers/programmers to emulate this approach; conditioned on the actual implementation environment. There are two main processing stages:

- **Evidence Retrieval:** where retrieved artifacts are mapped to an evidence record (ER) which is a meta-description used by downstream model operations and
- **Evidence Counting/Aggregation:** that describes methods used to transform event records into signaling evidence (artifact and social).

---

[107] Testing suggests that retrieval list composition is relatively insensitive to search depth setting over the top 25 or so ranked positions. However, while composition is stable, the rank order may vary across Low to High settings. However, retrieval depth is critical to recall; e.g., when identifying the whole expertise network.

Following the process descriptions, methods used to attribute evidence sources to actors and to assign organizational status to prospective experts are discussed. This includes:

- **Evidence Allocation**: which discusses "rules" are used to distribute evidence to experts in cases where there is missing or multiple-expert attribution (e.g., multiple authors) that must be resolved through contextual analysis and

- **Role Status:** which discusses a simple algorithm for assigning expert (signaler) organizational/role status; in the expertise model, status is associated with "honest" signaling as discussed in Chapters 3 and 6.

The two main processes are discussed next.

7.3.1   Evidence Retrieval and Overall Process

Expert Locator retrieves artifacts relevant to a particular query, as described earlier in this chapter. This includes ListServ postings, formal publications, project descriptions, and various types of online Web pages. Retrieved artifacts are transformed into signal evidence using a two-step process. In the first step, an evidence record (ER) is generated for each retrieved artifact. An artifact "identifier" (e.g., URL) is parsed to identify directory location which maps to the associated activity space and actor. For example, a transfer folder artifact, such as a PowerPoint briefing, is stored at *http://mii.mitre.org/xxx/yyy/zzz/transfer_folder/employee_id*[108]. The *employee_id* links to LDAP directory services which contains demographic data such as *Employee Name, Home Department*, *Site Location*, etc. With that, the AS Subspace is determined by the employee's *Home Department*; for example *G60: Information Technology Department*. Therefore, the retrieved PowerPoint is mapped to ER: {*AS= Personal, AS Subspace= "G60", Actor= "John Smith"*} which is essentially artifact signal evidence. More generally, the primary (partially annotated) ER fields are shown in Figure 7-9.

---

[108] This is a partial representation for illustration only.
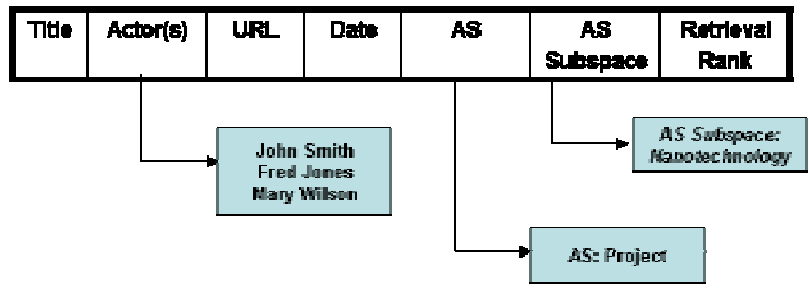
Figure 7-9: Core Evidence Record

The second step involves transforming an ER into signal evidence (i.e., counts) for each person across each Activity Space. ER's are used to populate the artifact and social evidence tables; that is, ER's are mapped to cell counts in the appropriate (artifact or social) evidence table as characterized in Figure 7-10; and described below.
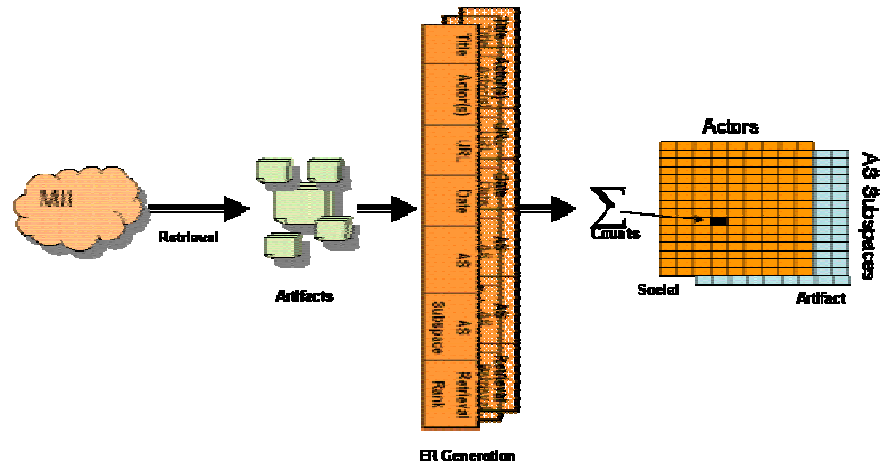
Figure 7-10: Overall Process

7.3.2   *Evidence Counting/Aggregation*

Table 7-2, below, provides context for evidence counting and aggregation described below. The table lists key sources used to generate ER's for each activity space (Projects, Personal, and ListServs).

AS	Sources	Artifacts	Source Motivation and ER Generation
ListServ	Specific ListServs	Postings	ListServ postings are artifacts generated from discussion group activity. Relevant postings are treated as artifact signal evidence; and the poster is as a membership instance, which in the aggregate across all Posters provides social signal evidence as discussed below. ListServ message headers are parsed to extract author, date, ListServ name, and other key fields used to instantiate the Evidence Record (Figure 7-9). Note: since ListServ threads are not exploited, the *Receiver* field in the post header record is not currently included in the ER.
Personal	Transfer Folders; Formal Publications; About-Me	Publications (Formal Publications, Transfer Folder items, and About-Me pages, and other, potentially, other artifacts for which there is text annotation).	Sources used in the Personal AS are not associated with discussion forums or formal projects. Persona AS artifacts include formal publications, transfer folder items, and About-Me pages all of which can be associated with a particular actor. As represented in the transfer folder example, above, artifact metadata is used to associate artifacts with a particular actor and actor's home department which establishes the Actor, AS, and AS Subspace fields in the ER record. Actor (Owner) is used as a membership instance which in the aggregate across all artifacts provides social signal evidence as discussed below.
Project	Project-Pages	Publications (Documents, briefings, and other artifacts having text descriptions).	Project documents are artifact signal evidence associated with a particular project. They represent work output associated with a particular activity; metadata extracted from documents is used to attribute ownership (i.e., authorship); however, where attribution is not directly determined, labor-level based rules are used to discriminate between "key" project personnel and peripheral staff. Key staff Poster is used as a membership instance which in the aggregate across all Posts provides social signal evidence as discussed below. Key staff are attributed to otherwise unattributed source evidence. This is discussed below.

Table 7-2: Evidence Record Generation

7.3.3   Artifact Evidence

Evidence Records are transformed into signal values and assigned to Artifact Evidence Tables[109] where table cell (*i,j, p*) contains the relevance-scaled artifact counts associated with (*AS_1, Subspace$_j$, Actor$_p$*); for example, *(AS_1=ListServs, Subspace$_j$ =BioTech, Actor$_p$=John Smith)*.

[109] The reader is referred to the formal model description, Chapter 6.

Simple ER counts are relevance-scaled to reflect signal value. For example, a single post in the Biotech ListServ with relevance rank three (3) has the cell value $(1/3)^{1/2}$; relevance ranking is obtained from the *Expert Locator* retrieval. Cumulative evidence for a particular subspace and actor $= \sum \sqrt[m]{1/k_i} \; \forall i$; where k is the rank of item i, and m is from [0, 1]. System default is $m=1/2$. Therefore, Artifact Evidence Table cell $(i,j, p) = |(AS_1=ListServs, Subspace_j =BioTech, Actor_p=John Smith)|= (1/3)^{1/2;}$ in the current example. Total signal strength for a particular actor in an AS Subspace is the sum of transformed inverse-rank weights. This is computed across all subspaces within an AS and is the $E_{i,\bullet,1, p}$ contribution to total evidence $E_{i,\bullet,\bullet, p}$ in Equation (6-5). This computation is repeated for all actors.

A representative Artifact Evidence Table is shown in Figure 7-11; the count data are contrived. The columns list candidate experts and the rows list AS Subspaces for each of the three activity spaces. For example, the first expert (Costa, Man) has personal evidence = "4", in the Personal Subspace (G026). The Social Evidence Table takes a similar form and is described below.

Figure 7-11: Representative Artifact Evidence Table

7.3.4   Social Evidence

Evidence Records capture social membership; for example, an author having artifact evidence in a particular AS, has membership = 1 in that AS. With that each expert is linked to one or more AS Subspaces; which in the aggregate, is viewed as a bipartite graph. For example, the ListServ

AS would form a bipartite graph with vertex set V{experts, ListServs} and with arcs {e}. A transpose of this graph is a 1-mode graph with V{experts} and edge set {e}; that is, experts are linked to experts they have common ListServ membership with. Then betweeness centrality, Wasserman and Faust (1994), is computed as a measure of nodal importance since it measures the centrality of a particular expert in the expertise community formed on the query topic. Figure 7-12 provides a representative expert-expert graph generated from the ListServ AS; node size is proportional to betweeness centrality. Betweeness centrality is the social signal value stored in the Social Evidence Table[110].
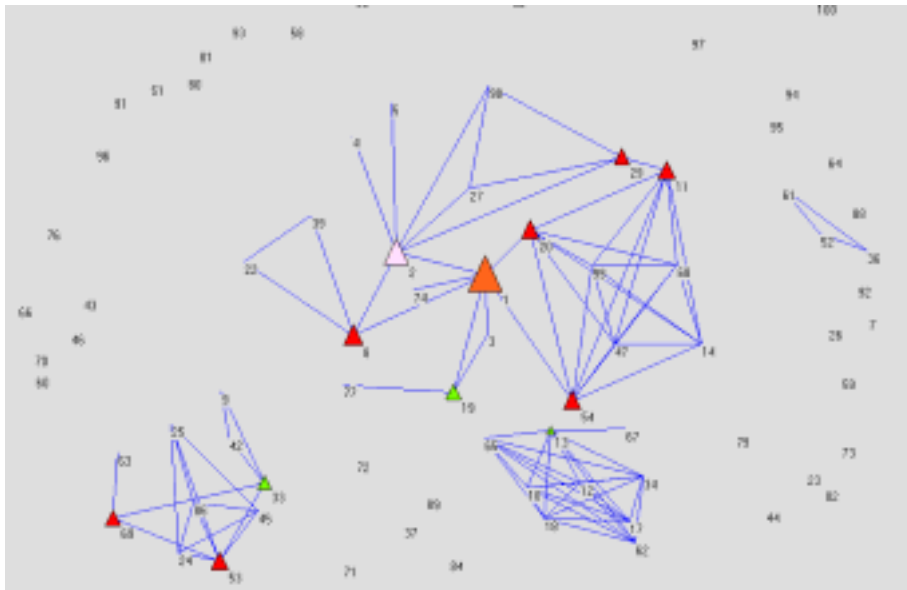
Figure 7-12: ListServ AS. Nodes (experts) Sized by Betweeness Centrality

7.3.5   Additional Model Computation

To complete the model computation the following steps are performed using the Artifact and Social Evidence Tables.

- Sum artifact and social evidence ($E_{i,•,•,p}$) for each actor and convert to Borda counts. This is $B(E_{i,•,•,p})$ in Equation (6-5).

---

[110] Note: In sparse social spaces, candidate experts behave as (near-) isolates—disconnected from most others. In that case betweeness centrality is not "unstable" and a simple degree measure (the number of links to others)s is used instead. Then social evidence score is the average degree computed for each expert across all AS Subspaces. This is done in each AS.

- Compute final actor score as the weighted average score across all AS's; scaled by actor's role status, R_p in Equation (6-5), and evidence distribution, N^γ in Equation (6-5).

- Compute actor expertise ranking based according to Equation 6-5; $\quad I_R(p \mid q) = R_p \bullet$

  $N^\gamma \bullet \sum \alpha_i \bullet B(E_{i,\bullet,\bullet,p})$. Note for N^γ , $\gamma = 1$.

7.3.6   Role Status

Actor status, used above, is derived from organizational role; with regard to signaling theory, Chapter 3, it is a proxy for signaler quality. The actors were partitioned into role-based classes: *Administrative/Support*, *Professional Staff*, and *Executive/Management*. The Administrative/Support category consists of administrators and certain technicians that provide infrastructure and desktop support. Professional Staff consists of scientists, engineers, and technical managers. Executive/Management consists of senior managers involved in day-to-day operations management and strategic planning.

Consistent with the emphasis on *technical* expertise, low weights (typically, $R_p < 0.2$) are assigned to Administrative/Support and Executive/Management staff. Professional Staff map into a finer-grained, 6-level scale that parallels the Applied Capability[111] rating. In the current setup, Figure 7-13, the Administrative and Executive categories are given relatively low weight; 0.1 in each case. Within the Professional category, status scaling increases roughly linearly so that R_p values for the seven AC categories are: AC1 = 0.15, AC2 = 0.2, AC3 = 0.25, and AC4 through AC7 = 1.0; the maximum status. R_p values are informed by user discussions and corporate policy; jointly used to determine culturally-sensitive status ratings. In that regard AC1, AC2, and AC3 are viewed as "junior" staff; They are often new hires with less work experience than the more senior AC's.

[111] Each member of the technical staff falls into one of the seven Applied Capability (AC) categories. Staff members are rated annually as part of the enterprise-wide performance reviews.
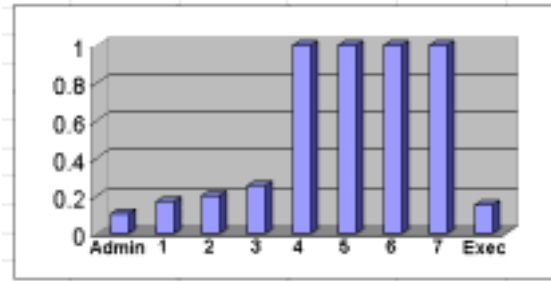
Figure 7-13: Sample Actor Status Ratings

7.3.7   Actor Attribution and Membership

There are instances where actor resolution (identifying an actor's identity and linking actor to artifact or social evidence) must be resolved in the absence of direct attribution.   Methods for handling special cases are discussed below.

7.3.7.1   Author Attribution

Artifacts are attributed to actors. This however raises a number of author resolution problems given the current enterprise publishing and document posting schemes. Published documents go through a standard metadata tagging operation in which authors are identified and attached to the document as separate metadata. ListServ postings have author identification built into the e-mail message header. However, other documents, such as those found in transfer (public share) folders, are not guaranteed to have been formally tagged or analyzed for authorship. In this case, authorship is problematic since the public share item may or may not have been authored by the share folder owner. Even when the owner and author are the same there may be co-authors not gleaned from simply assigning the owner as author. One approach is to use post-retrieval named entity extraction to extract authors from documents. This approach (rarely applied) does not scale well in dynamic retrieval environments where rapid retrieval performance is important and since adaptation of low-level indexing operations (via the Google corporate search engine) is prohibited, author resolution based on named entity extraction remains a longer-term development. As such, the default is to use share folder owner as author whenever artifacts were not formally published or for other reasons missing standard metadata.

Project artifacts present another instance of the author resolution problem since a project document space is analogous to a share folder. That is, project documents may lack specific author attributes in the same way share folders do. Therefore, when authorship is missing, the same scheme used to assign authors to share folders is used for projects with the exception that a project or task may be viewed as a multi-owner space. Depending on project structure (i.e., organization of tasks), task members are assigned as co-authors when other author attribution is lacking. With that, task members are treated as "equal" co-authors. A multi-author publication is parsed to reflect individual contributions; each contributor receives count = 1/n; where n= the number of authors. Note that projects (and tasks) may have large memberships; therefore a core membership is computed and forms the basis for assigning authorship. Membership filters are discussed next.

7.3.7.2   Project Membership Filters

Projects vary considerably in terms of the number of tasks and membership size. Project sizes vary along a continuum ranging from large sponsor projects to small internal studies (e.g., research tasking). Large sponsor projects are hierarchically structured with a project *root* or core task (usually associated with high level management functions) along with a number of tasks (*leaf* activities). Tasks usually consist of small teams typically having 10 or fewer core members. Whether a project has many tasks or none, the actual team may be arbitrarily large when actors who have minor roles are considered; this raises issues as to when and how project membership lists should be pruned[112]. The premise here is that core task members can be identified through analysis of labor expenditures. Of course, there are clearly issues with using task labor as a measure of "contribution" or role significance; however, the proposed approach works reasonable well in practice and is especially useful for removing likely "outlier" members; those who may only oversee a task or perform limited administrative functions. A two step approach is used to define task core membership. In practice, it is effective in eliminating actors that have peripheral roles.

---

[112] Control over task membership is important here as it affects social evidence measures sensitive to the size of a particular subgroup; for example a project..

Effort-based membership filters must address a built-in asymmetry in that task labor expenditures (as a proxy for productivity) may be normalized in two ways: based on total available actor labor over some period, or based on total task labor. Measuring effort along two dimensions stages for a type of portfolio analysis where an actor is a "core" member if contributing a significant percentage of overall task labor; or if expending a significant percentage of actor's available labor. This approach allows for special case handling; for example, on large projects an actor will typically account for a relatively small percentage of total effort even though he/she may be assigned full-time. So, both personal and task views on effort levels must be considered. With that, two "effort" measures are defined:

- a = actor task labor (hrs) /total task labor (hrs)
- b= actor task labor (hrs) /total actor labor (hrs)

Actor membership function, R, is defined as follows:

$$R = a*\alpha'+b*\beta' \qquad\qquad (7\text{-}1)$$

where, a and b are the effort ratios defined above, and α' and β' are normalized weights; so that, $\alpha'+\beta'=1$. The ranking function can be biased to selecting candidates with various work profiles; however, it practice it is defined so as to emphasize those that are heavily applied on the target task. Finally, actors are ranked according to R and the task membership list is cutoff at N_c, a system defined threshold; the default is 10.

7.4   Model Weighting

In the current model, weighting schemes are used in three main areas:
- **_Activity Space_** weighting is used to differentiate activity space importance as part of the merged ranking process. For example, Project evidence may be given more weight than ListServ evidence. As represented by α_i in Equation 6-5.
- **_Subspace_** weighting assigns relative weights of importance to particular instances of activity spaces. As an example, the ListServ space is made up of actual

151

enterprise discussion lists (i.e., subspaces); subspace weighting assigns weights of importance to each discussion list relative to a query. See β_{ij} in Equation 6-2.

- ***Evidence*** weighting is used to reflect the importance of particular kinds of evidence; in the current model relative weights can be assigned to *artifact* and *social* evidence. Refer to ω_{ik} in Equation 6-2.

Evidence, subspace, and activity space weighting schemes are discussed in greater detail, next.

7.4.1 Evidence Weighting

From Equation 6-2, ω_{ik}, is the weight assigned to evidence types within activity spaces. Currently, the evidence taxonomy is limited to two types: artifact and social[113]. The current model gives equal weight to each type; however, there may be a basis for weighting one type of evidence over another. For example, an individual's productivity, separate from that of, say, her connections to a group, may be most important and, in that case, artifact weight may be set higher than social weight.

The advanced retrieval interface allows users to adjust artifact and social weights as part of query generation. While preliminary studies suggest that minor deviations from, say, uniform weighting, have little affect on system performance there is evidence that the weightings can be used to cull out certain types of expert behavior. For example, increasing the social weight (relative to artifact weight) may be useful in identifying experts who were heavily embedded within a query-relevant work context but who had few artifacts. This may occur for experts new to the organization or project area for example; or it may suggest sparse artifact spaces that are social dense. In a practical setting, these "experts" might be given lower priority when selecting experts for independent work as opposed to collaborative tasking. In an opposite case, reducing the social weight elevates the relative importance of artifact evidence which may be useful in culling out high productivity individuals working in isolation. Identifying isolates that had high productivity may have special utility in identifying actors who may not be well integrated into

---

[113] Additional evidence types may increase model fidelity. This may include simply adding new types or in partitioning, say, social evidence into a finer grained categorization since there are wide ranging social contexts that may be usefully distinguished. In particular, social evidence may be partitioned so as to reflect formal and informal work.

core work or who may be assuming special roles. These characteristics motivate further investigation into adaptive evidence weighting strategies for ferreting out certain actor types.

7.4.2   *Subspace Weighting*

When aggregating evidence across an activity space; say, across a number of actual projects in the Project space, there may be a basis for assigning a higher weight to one subspace (project) over another. While several weighting methods were considered, the absence of training data motivated a simple uniform weighting scheme. While this does not preclude weighting certain subspaces higher than others, for example internal research projects, there was no clear basis for using specialized weights based on historical data or on query characteristics.

There is motivation to pursue specialized weights in future work as there were a number of instances in which non-uniform weights produced higher retrieval precision. For example, certain ListServ discussion groups have special importance with regard to a particular technology domain. For example, when searching for expertise in *link analysis*, the Analysis Cell List is the "richest" subspace to extract relevant postings and threads in that that list is populated by staff with expertise in developing or deploying analytical tools (such as link analysis). Similarly, for the Project activity space, internal research projects are especially useful in culling out experts in niche areas, and there is a relatively simple basis for assigning higher weights to internal research projects based on their internal project codes and domain classification.   While selective subspace weighting may be addressed through manual settings, it is problematic for users to adjust weights across large numbers of subspaces. Clearly, if user controlled weighting is to be used effectively, a suitable user interface is needed and this was outside the scope of the current prototype.

7.4.3   *Activity Space (Fusion) Weighting*

The literature on evidence combination includes significant work in information retrieval; for example, Aslam and Mantague (2001). From the perspective of the current work, evidence combination (fusion) strategies may be partitioned in terms of whether inputs are relevance scores or ranks, and whether training data is used or not. As described above, the current fusion

model assumes that each decision agent provides rank aggregates, and that mitigates problems with having to normalize scores from agent-specific score distributions Manmatha, Rath, and Feng (2001). In addition, the development of a large training set for optimizing fusion weights or other system parameters is problematic in the current environment. In particular, the process of generating relevance baselines does not scale well to expert finding where relevance judgments require judges with significant domain expertise. Therefore, the focus is on methods that do not require training data; machine learning methods are left for future work.

Several fusion strategies were explored to include uniform and manual weighting schemes. Uniform weighting is attractive given the inherent robustness of linear models, no need for training data, and ease of implementation. Alternatively, domain knowledge may be used to set weights manually and may be used as part of iterative search or relevance feedback. A third approach explored the use of a dynamic weighting scheme used to modify fusion weights on a query-by-query basis.

In the first instance, dynamic weighting suggests some kind of profiling method that captures the relevance of a particular activity space with regard to an expertise query. This is similar to collection profiling in distributed information retrieval systems that compute a probability of relevance for each collection and a given query; for example, the CORI algorithm Callan et al., (1995). Activity space profiling (i.e., weighting) may be feasible here, in the fashion of Balog and de Rijke (2007), if there is a reasonable basis for computing profile relevance on an a priori specified query or topic basis. However, topic detection as the basis for AS profiling is potentially a costly and complex operation not necessarily guaranteed to ensure sufficient topic coverage especially with regard to handling high-specificity queries related to emerging themes. Overall, the process of generating expertise topics and maintaining them over time is exacerbated by the cost to compute and continually update AS profiles across a large number of subspaces (e.g., thousands of ListServs) which makes this approach less attractive from a system maintenance perspective.

Instead, the approach explored here assigns fusion weights to AS decision agents based on their classification "behavior". In short, agent utility is related to *informativeness* (i.e., the amount of

information an agent provides about the relevance space.) Of particular importance here is the amount of information each agent provides when juxtaposed to the union of all other agents. This can be setup as a binary classification problem.

An agent's voting behavior is modeled as a binary decision process or channel where the outcome is either a one or a zero. Let X be a random variable such that

$$X = 0 \text{ (a candidate expert is not ranked by an agent)}$$
$$1 \text{ (a candidate expert is ranked by an agent)}$$

Then, the probability that any particular candidate is ranked by decision agent A_i is (i.e., the probability that X = 1):

$$p_i = n_i / N \qquad (7\text{-}2)$$

Where n_i is the number of candidates ranked by agent A_i, and N is the total number of unique candidates across all agents. Then, the entropy (information) associated with agent A_i is:

$$H(x_i) = -\sum p_i * \log(p_i) \qquad (7\text{-}3)$$

and for, X, a binary random variable, the binary entropy is:

$$H(x_i) = -p_i * \log_2(p_i) - (1 - p_i) * \log_2(1 - p_i) \qquad (7\text{-}4)$$

Equation 7-3 is repeated for each agent, A, so that the normalized weight for each agent is:

$$W(A_i) = H(A_i) / \sum_{\forall i} H(A_i) \qquad (7\text{-}5)$$

Then, $W(A_i)$ is used in Equation 6-4 where

$$a_i = W(A_i) \qquad (7\text{-}6)$$

8   **Evaluation Issues**

In one sense the enterprise is a "hostile" environment in which to conduct an evaluation; there is a lack of experimental control compounded by operational constraints imposed by the host organization. Here, there was no existing system to compare *Expert Locator* to, no training data to baseline the new system against, and no a priori knowledge of what constituted relevance for a given topic—inhibiting the development of a test collection. The remainder of this chapter discusses how operational constraints factored into the evaluation in areas such as: test query generation, relevance assessment, and results scoring. While the evaluation model used borrows from large-scale evaluations like TREC, the evaluation of *expertise* relevance as opposed to *document* relevance required a new approach to building a test collection and to assigning relevance to <u>people</u> and not <u>documents</u>.

8.1   Evaluation Design Issues

Information retrieval evaluation is central to the development of new search technologies and working systems. Early work in evaluation, for example, Salton and McGill (1983), focused on small collections which made it feasible to assess document relevance over the entire collection. However, the need to scale-up retrieval system performance to handle massive data sets, to work in mixed language environments, and on novel retrieval applications has continued to motivate new evaluation research. The DARPA initiated <u>T</u>ext <u>RE</u>trieval <u>C</u>onference (TREC), Harmon (1993), has been instrumental in developing scalable evaluation methodology to work across a number of large-scale search tasks such as web searching and question answering. TREC and other large-scale evaluation efforts address a number of scalability issues related to data collection, relevance assessment, and performance measurement. A number of these issues are common to expert finder evaluations.

Table 8-1 outlines some of the more significant issues that cross-cut IR and expert finder evaluation and how they are addressed in this research. This is followed by a more in-depth

discussion. The emphasis here is on evaluation issues that discriminate operational expert finding evaluation from traditional IR assessments.

IR Evaluation	Large Collection Issues	Expert Finding Evaluation
Generally a static collection	Collection	Operational Environment
Documents	Objects	Actors (Experts)
Relevance Pooling; Non-pooling Approaches	Collection Relevance Assessments	Snowball Sampling— Consensus Ratings
Binary; Multi-level	Relevance Levels	Multi-category
Missing Document Relevance Information	Completeness	Missing Expert Ratings
MAP, R-precision, MRR, P(@10),…	Performance Measures	R-precision, P(@5), Awareness,…

Table 8-1: Large Collection Issues

8.1.1   Collection Environment

Operational environments add complexity to system use and evaluation. As such, *Expert Locator* is inherently more complex than a "laboratory" retrieval capability in that it had to integrate with Intranet services and work in concert with corporate policies addressing information access and security. For example, in some cases Intranet data was only obtainable dynamically on a per query basis; for example, from Directory Services. In other cases, the evaluation system mirrored a corporate collection to facilitate real-time data capture and effective access; this was the case for ListServs where *Expert Locator* performed real-time capture of daily postings, maintained a separate ListServ database, and indexed postings for retrieval. Added to this, *Expert Locator* could only monitor ListServs for which the "owner" agreed to have postings re-hosted. While ListServs are public there was a "privacy" concern regarding pooling multiple postings for the purpose of identifying usage patterns. Each ListServ

owner had to be polled, and where access was granted a special *Expert Locator* "user" was set up to receive postings.

To enhance system stability for testing, a special version of the system was set up so as to minimize real-time access to Intranet services and to shield users running the standard prototype from evaluation activities. It was especially important that system usage and evaluation procedures not adversely affect network loading, users' work activities, and mission performance. Most evaluation processes were run in the off hours to minimize resource contention. While most of the special set ups and processes run do not impact retrieval accuracy they do increase the complexity of the overall assessment.

There are other operational test issues that may impact the stability and accuracy of various experiments. Since the Intranet (services, collection, user interactions) is changing over time it is important to restrict testing to as short a period as feasible to reduce the impact of changes in the underlying information space and user interactions. For example, if the evaluation was run over a period of months, the actor pool may change substantially (e.g., new employees), roles could change, and ListServ traffic or publications could exhibit major topical shifts. This could affect expertise ratings for some queries. Therefore queries were processed in roughly 1 day and the results archived for analysis. This "snapshot" was policy-restricted to contain expertise rankings and related experiment parameters only; corporate policy precluded archiving the entire Intranet representing roughly 4000 users, more than 10 million artifacts, several hundred organizations, and thousands of project tasks. The analyzed information space represents corporate work performed over more than five years; although the distribution of artifacts across work forums is not uniform since project spaces, ListServs, and various personal data spaces were not instantiated all at the same time.

Just as a particular test collection, say a news source, may not be complete in terms of covering all news stories, enterprise data exists in enclaves and may be inaccessible to collection and analysis; as such, missing evidence may impact whether *Expert Locator* will judge an actor as having certain expertise or not. There are two cases of interest. First, since the current system is bootstrapped on three activity spaces initially, there may be activity spaces missing that could

substantially effect performance and, depending on the query, the system may under perform. In the second case, information may be compartmented based on privacy or security classification. This problem is more difficult to address; however, to the extent those cases are mirrored by public data, the system may still perform well. Absent that the system will not reflect a particular expertise area; this is more likely to occur for sensitive problem areas as opposed to general technology domains.

While evaluations conducted in operational environments present unique challenges compared to "laboratory" assessments, there are areas of common concern. In particular, in both instances relevance judgments performed on large collections across a range of queries are costly and require formal procedures to ensure reasonable collection coverage. As such, both IR and expert finding have complex relevance landscapes which must be navigated by expert judges in one fashion or another in order to ferret out query relevant sets, *qrels*. Scaling relevance assessments to large collections is clearly a problem inherent to both expert finding and IR.

8.1.2   Evaluation Objects

A central divergence between large-scale IR evaluations like TREC and expert finder assessments is the notion that the target relevance set is made up of people and not documents. Essentially, relevance assessments must address a number of issues that separate experts from documents as retrieval objects. This is addressed in Table 8-2, below, where documents and experts are embedded in their respective evaluation paradigms. From there, contrasts between traditional information retrieval and expert finding stand out.

Documents are evaluation objects in information retrieval while expert finder evaluation must address artifact evidence (propositional) and relational evidence such as links between experts and activities. Both evidentiary sources may be useful in assessing relevance. For example, if a person has significant credentials in some area, that person may be judged an expert based solely on propositional evidence; e.g., number of papers published. However, if the same person is linked to known experts or recommended by them then relational evidence (context) provides

another basis for assessing relevance. Therefore relevance is contextualized through an experts' embeddedness in a relevant work context.

Relevance may be based on self-ratings, peer ratings, referrals, or affiliations if available; clearly not applicable to documents. This is relevant in that expertise assessments made by peers as part of normal work or system usage (i.e., gathered through relevance feedback) may be used to assess actor expertise level on future queries. Documents are "passive" and are generally treated as simple artifacts. Experts, however, can self-organize and form groups or communities within which they may take on explicit roles such as broker or practitioner.

Aspect	Information Retrieval Evaluation	Expert Finder Evaluation
Objects Assessed	Documents and potentially relations between documents and authors; although this is not typically performed in large scale evaluations such as found in TREC.	People and indirectly Propositional (documents, activities, events, location, etc.) and Relational (e.g., associations)
Object *awareness*	None -- documents lack *awareness*	Experts self-rate
Object Groupings	None -- evaluation does not typically factor in document groupings within a collection	Affinity groups, communities of practice
Object Linkages	None --evaluation does not typically factor in document linkages or inter-document ties	Peer-to-peer ratings
Object Roles	None -- documents are simple artifacts; they are not assigned functional roles	Multiple roles (e.g., broker)

Table 8-2:   Evaluation Objects

8.1.3   Collection

The use of large IR test collections has exacerbated the problem of assigning relevance to documents.   Selecting qrels is problematic and does not lend itself to standard sampling approaches. This has led to the use of document pooling as the de facto approach for generating relevance judgments for large collections; which raises issues as to the efficacy of pooling and related approaches to expert finder evaluation.

A simplistic view of document pooling has the first *k* items from multiple retrieval systems pooled to form an initial nomination set. Then, pre-selected judges assess pooled items and generate qrels to be used in the evaluation.   With this approach, the high cost to manually review candidate documents is mitigated somewhat by reducing the raw pool down to the items that are

system nominated. Pooling techniques of this sort are widely used today, for example, TREC, Voorhees (2003), and CLEF, Peters and Borri (2004); with earlier work tracing back to Spark Jones and Van Rijsbergen (1976) amongst others.

Modifications to standard document pooling have been studied extensively. For example, the Move-to-Front (MTF) method, Cormack et al (1998), modifies the TREC approach of treating all contributing systems alike by adjusting the number of items each contributes based on retrieval performance. Systems that have a higher probability of relevance are weighted higher and will necessarily submit more documents than lower performing systems[114]. Using weighted nominations, Cormack et al found it possible to reduce by ½ the number of relevance judgments needed to generate effective qrels[115].

Cormack et al (1998) also used iterative searching to generate qrels. Their method, Iterated Searching and Judgment, (ISJ), uses query reformulation as a basis for generating a relevance set. Essentially, judges interactively search a collection for some (arbitrary) period in order to locate relevant documents. Searchers typically reformulated the query at each stage or terminated the search depending on the quality of results received. The process was effective in that less than ¼ as many judgments were needed to generate effective qrels. Soboroff, Nicholas, and Cahan (2001) took a more radical approach by exploring various ways to generate a raw document pool for use as qrels and then assigning relevance assessments randomly. However, while the approach discriminated medium performing systems from poor ones, it was not useful for discriminating between the best and worst systems. Regardless of the pooling method, it is difficult to make a strong case for using these techniques in an *Expert Locator* operational assessment. In the target environment, multiple systems are not available for document pooling and expert judges are costly and difficult to assemble on a query-by-query basis.

Sanderson and Joho (2004) assessed ISJ for use with a single system. They found evidence that a single system (regardless of relevance feedback strategy tested) can generate usable qrels. More

---

[114] The approach is similar to collection weighting schemes for heterogeneous retrieval Voorhees et al (1995).

[115] Here effectiveness refers to the correlation between system rankings; for example, if Kendall (MTF, TREC) > 0.90 the two pooling methods generated the same system rankings.

specifically, three systems were used to generate qrels using the modified ISJ approach and compared against TREC. At each stage, the query is modified using a particular relevance feedback scheme. Rank correlation (Kendall's Tau) is used at each iteration to quantify rank order similarity between the ISJ/Relevance Feedback approach and the TREC baseline. The correlations improved with successive iterations; although there was no evidence that the relevance feedback method used was a significant factor. There was, however, some indication of system variation; one system had, on average, higher correlations (0.93) than either of the other two (0.87 and 0.89); this may suggest further study. Overall, the authors concluded that when using relevance feedback, modeled after the approach used by Soboroff and Robertson (2003), system pooling was not needed to generate effective qrels. In additional experiments, Sanderson and Joho found evidence that non-pooling methods could produce usable qrels. This was based on using ISJ to produce qrels from manual and automatic runs. Systems were ranked using mean average precision and correlated with results from four different TREC evaluations. Using the *Voorhees acceptance level* (correlations > 0.8 are significant), 88% of the manual runs and 77% of the automatic runs produced *usable* qrels. The results are surprising, especially for the automatic runs, and they have implications for future large-scale IR assessment and for operational tests where multi-system comparisons may not be feasible.

The approach taken here departs from the non-pooling ISJ approach which *centralizes* relevance judgments to one or a few a priori defined raters. To build organizational consensus on expertise ratings, the supposition here is that the evaluation requires a distributed, multi-rater scheme in which the raters are drawn from the same pool as the actors being rated; in other words *using experts to rate experts*. This "circularity" is addressed in part by a survey-based *voting* scheme; however, one in which voters are not pre-registered, but are identified dynamically as part of the voting process. To a large extent this fits snowball sampling, Snijders (1992) and Berg (1988), which, in this case, is used to generate a consensus graph encapsulating expertise ratings and knowledge of who-knows-whom through a single process.

Snowball sampling is similar to web crawling, Konchady and D'Amore (2002), in which given some seed pages, page links are traversed to identify a progressively larger data set. A key difference here is that where web crawlers typically use a single rule set to judge page relevance

and navigational control (links to follow), snowball sampling distributes decision making to each new generation of nodes evaluated. Therefore the snowball process can be viewed as a form of behavioral "averaging" across a set of voters each with their own intrinsic rules and (private) knowledge as to what constitutes expertise and who qualifies as an expert. Organizational consensus on expertise ratings is established based on the distribution of votes across experts identified and this scheme can be used to establish trust in the evaluation and in an operational system. With that, the Expert Locator evaluation is contrasted to traditional IR and prior expert finder evaluation methodology as shown in Table 8-3.

Evaluation Basis	IR-TREC	Expert Finding Literature	Expert Locator Evaluation
Relevance Sets (qrels)	qrels (based on pooling)	Typically post-retrieval Assessments	Snowball-generated Query Relevance Sets (s-qrels)
Relevance Judgments	Judges	Judges/Panels	Self-ratings and Peers
Roles	None	None	Multiple (practitioner, broker)
Sensitivity Analysis	Various indexing strategies, etc.	Not typically addressed	Evidence Combination assessed

Table 8-3:   Relevance Handling: Different Approaches

As noted in Table 8-3, there are several areas where the *Expert Locator* evaluation diverges from prior efforts. For example, the evaluation conducted here uses actual experts from the target environment to assess candidate experts and experts polled are "selected" consistent with target queries. This is derived directly from the snowball sampling scheme. The evaluation also supports system performance assessments as a function of expert's role; this will be introduced later in this Chapter and in more detail in Chapter 9. Finally, the approach taken here assesses system robustness with regard to variation in the sources of evidence used. In particular, experiments are run that assess how the number of activity spaces used affects precision; for example, are two activity spaces always better than one? The author is not aware of any expert finding evaluation in which detection rates are calibrated as a function of the type of evidence used or where roles have been "computed" based on network position. As such, the snowball sampling scheme seems reasonably well suited for expert finder evaluation, supporting new kinds of assessments.

8.1.4   Missing Information and Performance Measures

Snowball sampling has distinct advantages over random sampling when working with hard to detect or sparse subpopulations. It is a name generator that ferrets out nodes and edges in a graph through a type of referral process. However, snowball sampling is biased[116] and convergence on the relevant population is dependent on network structure, initial sample points, and resource constraints. Snowball sampling has behavioral similarities to diffusion or disease transmission implying that organizational network structure may either inhibit or promote edge formation (i.e., survey responses). Given certain initial sampling points, the "spreading activation" behavior of the snowball can lead to dead ends; which results in missing relevance judgments. As such, experts missing from the snowball sample are equivalent to the problem of *missing information*, endemic in large-scale retrieval evaluation.

Buckley and Voorhees (2004) assessed the impact of missing information on performance measures. They found that traditional measures, such as P(@10), R-precision, and mean average precision (MAP) are unstable with high levels of incompleteness. They advocate a new measure, *bpref*, which they found to be fairly robust to incompleteness. In their experiments, as the number of relevant items decreased (by removing items), system rankings using bpref correlated well with rankings using complete qrels while measures like MAP and R-precision degraded (especially with 50% incompleteness or more.)

The bpref measure essentially counts the number of known non-relevant items that are ranked ahead of known relevant, when performed over R ranks. Unrated items are ignored. However, if there are many unrated items and few non-relevant items, bpref is less useful. This is relevant to *Expert Locator* evaluation since accounting for missing information (unknowns) is important in two regards, first as an indication of snowball sample coverage, and, second, as it reflects on whether the system is finding experts not visible to the average expertise network member.   This suggests modifications to bpref or possibly different measures need to be used.

---

[116] Recognizing that document or query pooling methods are also biased since the "sample" is generated by one or more systems using a non-random selection scheme.

In certain usages at least, *Expert Locator* assessment is more closely aligned with Question Answering (QA) evaluation where the focus is on finding a few key items (experts) with the goal of attaining high precision; for example, Voorhees and Tice (2000). QA is a special case of high accuracy retrieval; however, Shah and Croft (2004) noted that precision and recall are generally unsuited for measuring performance in high accuracy retrieval applications. They advocated the mean reciprocal rank (MRR) <u>of the first relevant result</u>. While this may be suitable for select cases (e.g., the TREC High Accuracy Retrieval from Documents—HARD[117]), it is rather restrictive for expert finding where system effectiveness is related to providing users choices of which experts to contact. *Expert Locator* should be evaluated on more than just the top ranked retrieval, the position of the first relevant item, or other measures that mask the systems ability to provide a high-precision "short list". To amplify this point, the system may be most useful when it retrieves clusters of experts where multiple clusters span a range of work or mission areas and organizations.

In summation, while Buckley and Voorhees' findings make a general case for using bpref, the discussion above argues that it may not be as suitable in cases where *unknowns* matter as in the *Expert Locator* evaluation. In addition, measures that isolate performance to say the first relevant item are restrictive and less useful for conveying the *Expert Locator*'s utility in an operational environment. In Chapter 9, two performance measures will be addressed that on balance provide a reasonable basis for assessing performance.

8.2   Query Generation

Topic areas were generated based on inputs from various sources; however main emphasis was placed on topics aligned with key organizational technology areas. While this is a small subset of the technologies or problem areas of interest it is representative and covers main business areas and operating centers. A second source of topics consists of email requests broadcast to relatively large segments of the enterprise (since users don't have a good sense of who knows what in the niche areas they often broadcast queries to large groups). These queries tended to be

more specific and might include queries that focused on a particular application, a customer, or a particular information source. For example, a query might be "Does anyone have any knowledge of WebTas?" or "G60 needs some logistics help…" While informal use of *Expert Locator* on this type of query was encouraging, a formal analysis is reserved for future work. The queries used in this study are more general and are listed in Table 8-4, below.

Air Traffic Control	Bayesian Networks	Biocomputing	Biometrics	Brain Mapping	Chemical Warfare
Complex Adaptive Systems	Geospatial Mapping	Global Position System	Grid Computing	Homeland Security	Human Computer Interaction
Information Retrieval	Insider Threat	J2EE	Logistics	Nanotechnology	Network Protocols
Operations Research	Robotics	Satellite Communications	Semantic Web	Signal Processing	Simulation and Modeling
Social Network Analysis	Software Engineering	Speech Recognition	Vegetation Forensics	Wearable Computing	

Table 8-4: Evaluation Topics

8.3   Establishing a Relevance Baseline: The Survey and Snowball Sampling

A survey is a method used to gather information from a group of individuals[118]. It differs from a census in that a survey samples only a subset of the target population. A survey is usually based on random sampling; however, where the target subpopulation is unknown or sparse, simple random sampling may be inefficient. Here, snowball sampling is used to generate a sample and the process is initiated by identifying an initial seed group. For each query an initial group is nominated using various methods to include inputs from resource brokers, retrieval systems, and a priori known lists of relevant experts. Each respondent is sent an introductory email outlining the evaluation goals and requesting their participation. The email has a link to the survey form which can be filled out online.

---

[117] http://ciir.cs.umass.edu/research/hard/ Accessed on January 4 2005

[118] http://www.amstat.org/sections/srms/brochures/survwhat.html Accessed on 8 January 2005

The survey form, *Appendix: Survey Form*, was designed to be filled out quickly, and to be easily automated for use online use across a wide user base. Each respondent was asked to respond to several questions regarding their experience in the query domain as a way to establish a working context for obtaining rating scores, Foddy (1993), Converse and Presser (1986). User's were also asked to self-rate using a 5 point Likert scale; with level 1 associated with having little knowledge of the topic being surveyed. A sixth option "Not Sure" was added to identify any individuals that had problems with assessing their expertise level; however it did not affect the actual analysis.

Each survey recipient was also asked to assess eight other people; using a 5-point Likert scale. Unknown to the user, the eight people consisted of five individuals that were likely to be relevant to the query (excluding the recipient) and three other randomly selected from the general population (most likely non-experts). Users could also nominate names not represented in the list of eight candidates. As such, the form balanced out direct assessment (a roster) with recall-based nomination to provide some measure of coverage on the target population.

The actual experiment was conducted in two phases due in part to organizational constraints imposed on survey duration and also to mitigate work schedule conflicts. The first phase was conducted in July, 2003 and the second phase in September, 2003. In each phase, users were sent e-mail reminders prior to the final due date; consistent with research suggesting notification increases response rate by as much as 25%., Sheehan and Hoy (1997). Survey mailings and response rates are noted for each phase in Table 8-5, below. Pragmatic considerations imposed certain constraints on the survey process; for example, to ensure adherence to organizational "protocols", each expert was limited to 4 survey forms (unique queries) and no individual was asked to fill out the same survey more than once. A 41% response rate is generally accepted as being reasonable for an e-mail based survey[119].

Phase	Sent	Received	Rate
July	456	178	0.39
Sept	841	355	0.42
Totals	1297	533	0.41

Table 8-5: Survey Mailings and Responses by Phase

[119] http://www.ascusc.org/jcmc/vol6/issue2/sheehan.html Accessed on December 28 2004

Overall, 29 queries were run. The distribution of survey responses across queries is shown in Figure 8-1, below. The average response rate was 9.48 per query. The average size of a snowball query relevant set, s-qrels, is approximately 32. Therefore, the snowball generated roughly 3 times as many nominee ratings as there were actual survey responses. In this case one doesn't have to self-declare as an expert to be known organizationally as one. This adds a certain level of robustness to the survey collection.
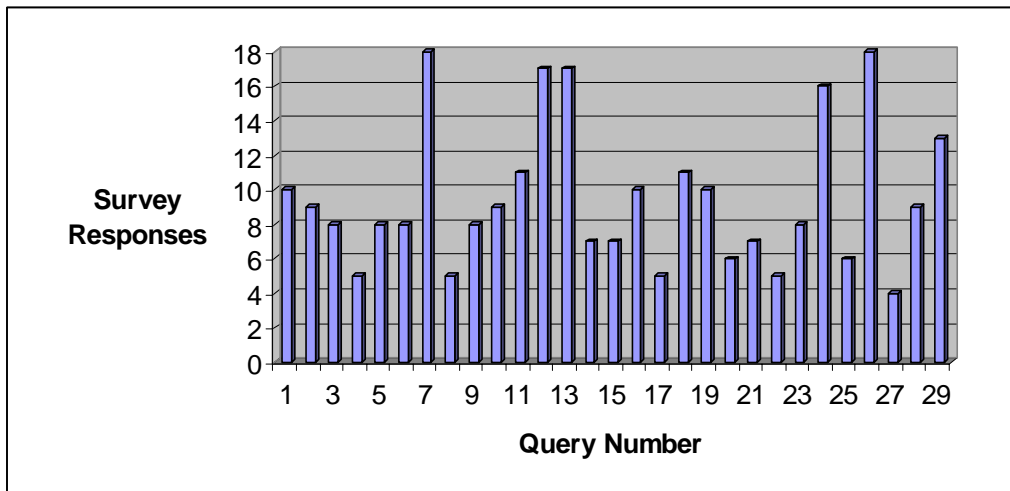
Figure 8-1:   Survey Responses per Query

8.4   Data Collection

Survey results were stored in a relational database for easy processing. The data for each person surveyed can be viewed as a 6-tuple [person, {person-attributes}, topic, self-rating, {peer-rating}, {nominations}]. This is rendered partially in Figure 8-2 as an ego-centric graph. Here the surveyed person is the *ego* at the center of the graph, and those rated are *alters*; this includes those that were *peer rated* or *nominated*. Each ego has organizational attributes such as home department, location, room number, mail address, technical level, as well as topic-dependent descriptors such as peer rating and nominations. In addition, each person can self-assign themselves a role such as *practitioner* or *broker*, both or neither. While the ego-centric graph shown in Figure 8-2 is a simplification, it can be viewed as a component of a more complex graph; one that describes the entire snowball sample.
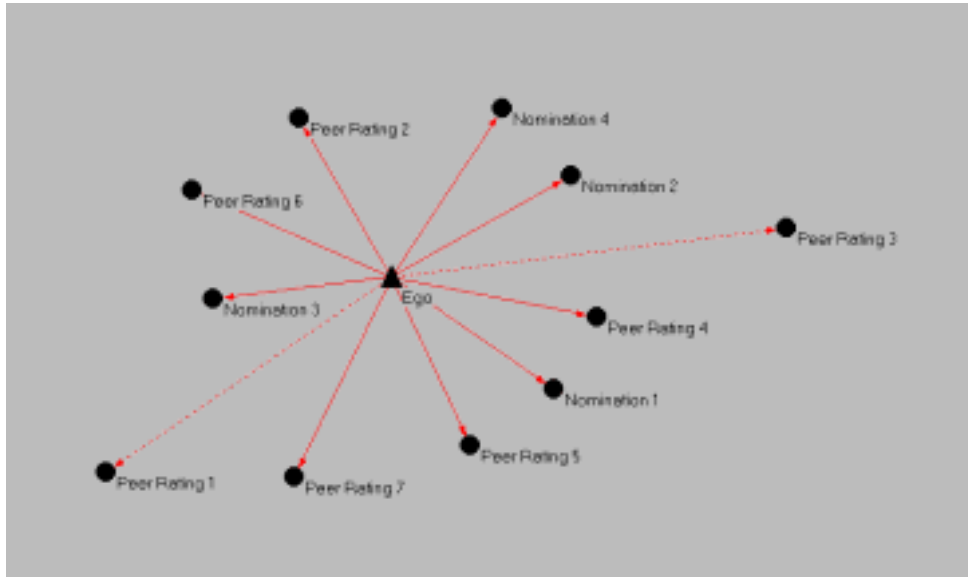
Figure 8-2:   Respondent Survey Graph for a Particular Topic

The topic graph shown in Figure 8-3 represents all the relationships identified through the snowball sampling scheme. Node labels are numbered to ensure anonymity while edge weights reflect the expertise rating assigned by the source (rater) to the sink (the person rated). Here, the peer rating scale used in the survey has been transformed so that it is now contained within the interval [-2, 2]. Now, peer ratings that reflect *disagreement* that a person is an expert receive negative scores [-2,-1], zero represents a rating of *uncertain*, and positive values [+1, +2] are associated with *agreement*. Using this scale negative ratings are represented with a broken edge line. Isolates represent individuals that did not receive peer ratings. Finally nominations are given a default rating (edge weight) of +1. Overall, the topic graph (snowball) reflects group consensus on who is an expert within the expertise network and, as discussed in Chapter 9, network structure is used to assign a relevance score to each person and to identify *roles*.
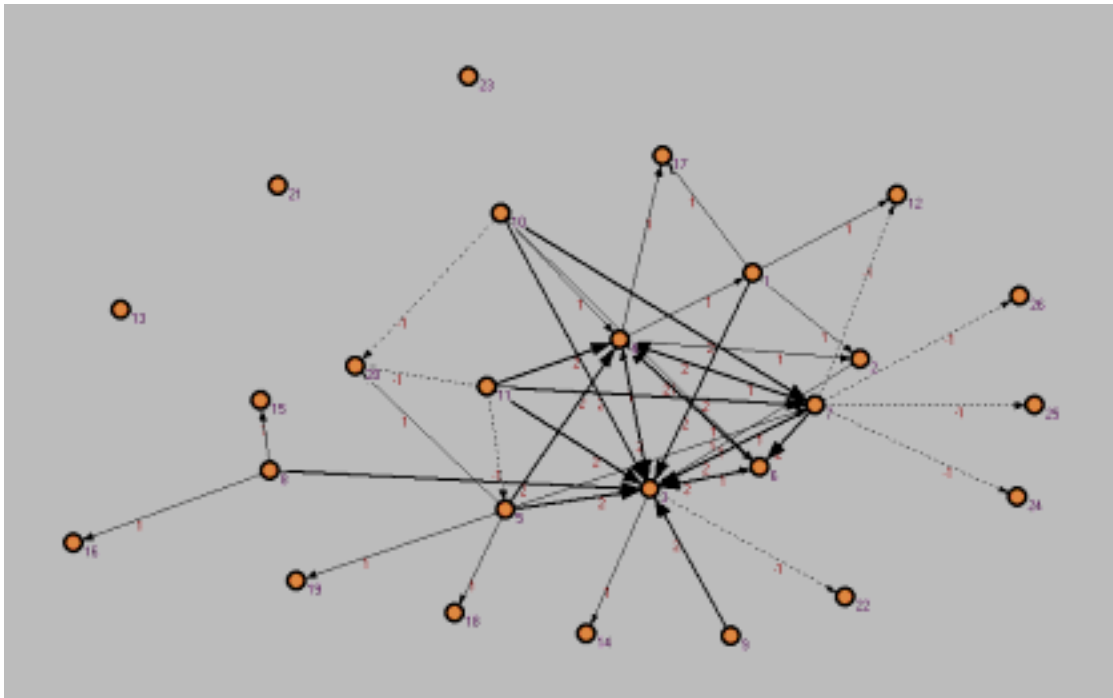
Figure 8-3: A Snowball-derived Consensus Graph for a Topic

The next chapter focuses on actual testing and results. Selected precision-based measures support a broad assessment of Expert Locator performance with regard to variation across queries; the effect of using evidentiary sources in various combinations—eliminating some, combining others; and, the role that missing information plays in the assessment.

9 Methodology and Results

The chapter covers experiments used to assess *Expert Locator* performance to include measures of system robustness to variation in queries and sources of evidence used. The chapter begins by developing a method for converting snowball graphs into s-qrels; essentially a list of actors relevant to an expertise topic area. This process sets the stage for the precision-based assessments that follow.

9.1 Establishing the Relevance Baseline (s-qrels)

The snowball-based survey described in Chapter 8 provides several bases for determining relevance and non-relevance. Actor self-ratings provide a direct assessment; however, this raises issues regarding the efficacy of using self-ratings to compute system precision. To address this peer ratings and nominations were used as a basis for validating self-ratings. For example, if a candidate self-rated as having expertise and if there was peer agreement, then it was assumed that the self-rating was valid or at least consistent with outside opinion. In an initial sample (n=167) from the total experiment, 107 (64%) self-ratings were peer reviewed. From this, the self-rating reliability was computed as a measure of consensus between self-ratings and peer review. Overall, self-rating reliability was 93%. That is if a actor self-rated with survey score greater than or equal to two (on the original 5-point Likert scale), and if that person's peers rated the user as having expertise either through nomination or peer rating then the self-rating was validated. Using this form of voting, in cases where the respondent self-rated as an "expert" the reliability was 95% and when the respondent self-rated as a "non-expert" the reliability was 89%. This is significant in the context of the queries evaluated here and provides some confidence in using self-ratings as indications of relevance. However, the situation degrades when considering the whole experiment. Then, only 31 percent of the surveyed group actually self-rated; as such even though self-ratings seem to align with a vote of one's peers, using self-ratings to compute precision would force discarding roughly 69% of the relevance information collected. As such, an alternate path is taken.

As an alternative to self-ratings one can appeal to the snowball graph for more complete relevance information. In particular, the snowball generates a graph in which candidate experts are represented by nodes and arcs that represent the level of expertise one candidate ascribes to another. This sets up as a voting scheme of sorts; votes received (in-links) are used to gauge consensus as to whether a candidate is an expert or authority on the topic. Votes submitted for others (out-links) are used to identify "brokers", i.e., those that have knowledge of true experts. Ideally, the voting model would converge on the best brokers and experts as well as juxtapose single class actors, such as brokers, with those playing both roles. This perspective on experts and brokers aligns nicely with the notion of hubs and authorities (HITS), Kleinberg (1999), in which nodal importance can be viewed in terms of a hub score (actors that point to the best authorities) or an authority score (actors that point to the best hubs). Here, hubs and authorities can be interpreted in the context of different expertise network roles without having to formally poll for such information. Essentially, everyone gets a hub and authority rating without having to self-rate through the formal survey; it is based on peer assessments.

Using HITS one can compute the hub and authority score for each candidate. Figure 9-1 depicts a snowball-induced graph with node size reflecting authority scores. Here, authority scores are used to rank nodes according to expertise level. Arc weights reflect peer ratings on the interval [-2, +2]; with negative scores counting as votes against a person having significant expertise. In the graph, positive ratings have solid lines and negative ratings have dashed lines. For this example, there are four candidates (large nodes) that received high authority scores with a number of others that receive lower ratings to include several that are negatively scored by their peers. A similar computation is done to compute hub scores and the ordination generalizes; composite ratings are generated from a simple linear additive model.
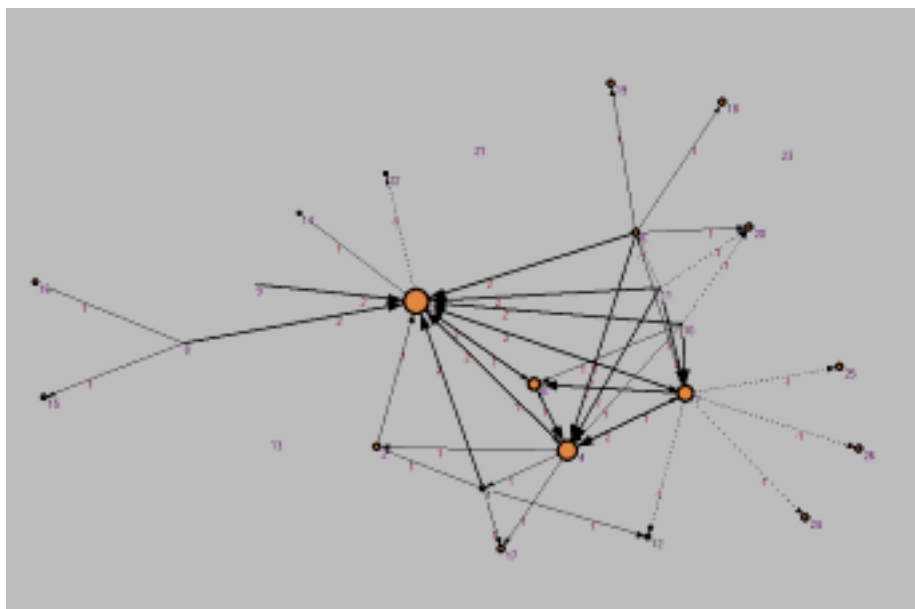
Figure 9-1: Graphical View of Authorities based on HITS Algorithm

The snowball sampling technique coupled with HITS provide a novel way in which to build consensus as to who is an expert or can point someone to an expert. Using this approach, a relevance baseline, s-qrels, is built for each query, and in the next section *Expert Locator* retrieval lists are compared to this baseline using two performance measures.

9.2   Introduction to Experiments

The evaluation is discussed in two parallel tracks, essentially. The main assessment follows a traditional IR evaluation in that system performance is based on precision measures augmented by system robustness assessments.   In parallel, selected evaluation questions are recast so as to address the underlying theory; in particular experimental findings are viewed in light of Signaling Theory and Activity Theory.   A caveat here is that given actual signaling behavior is likely more complex than the simplified model presented in Chapter 6, discussions linking precision results to signaling theory are purely exploratory at this point and, at a minimum, serve to motivate future investigation.

Beyond the current experiments, a short discussion on alternative evaluation measures is presented in Chapter 9.3; where the evaluation is shifted from precision-based to one that is focused on the amount of new information provided to a user. While the actual experiments are not carried out in the current work, evaluation design and performance measures are discussed. The intent is to motivate system performance assessments with respect to locating experts not previously known to a user.

9.2.1   Experiment 1: Overall Retrieval Performance

This section compares *Expert Locator* retrieval performance (i.e. detection) to actual organization experts. Based on snowball sampling consensus ranking, system retrieval lists are compared to snowball relevance graphs to assess overall retrieval performance from several perspectives.

9.2.1.1   Approach

An *Expert Locator* retrieval list consists of *known relevant*, *known non-relevant* and *unknown* items. Known relevant and known non-relevant are obtained from the snowball sample with the computed authority or hub scores used as weights. Unknown actors are not represented in the survey; so that the basic approach computes precision while treating the unknowns as falling into one of two classes; relevant or non-relevant. This provides a basis for computing an upper and lower bound on actual performance.   The relevant population is usually small, especially for sparsely populated expertise areas. Therefore, it is not generally feasible to compute precision at fixed points; R-precision is used instead. R-precision is computed as:

$$P(@\ r) = RELret\ /\ Rknown \qquad\qquad (9\text{-}1)$$

where *RELret* are the relevant items retrieved by *Expert Locator* and *Rknown* is the total relevant set (obtained from the snowball sample)[120]. The lower bound on *Rprecision* can be computed as shown in Equation (9-2).

$$L_{P(@\ r)} = RELret\ /\ Rknown \tag{9-2}$$

The lower bound computation is based on the assumption that all unknowns are non-relevant and are added in with the non-relevant retrieved. This is done by counting only the *RELret* in the numerator. The upper bound, Equation (9-3), adjusts the numerator used in the lower bound by assuming, in the best case, that all unknowns within the top r ranks are relevant.

$$U_{P(@\ r)} = (RELret + UNKret)\ /\ Rknown \tag{9-3}$$

From an expertise network perspective, R-precision can be used to assess what proportion of the expertise network is retrieved in the first R ranks. A more restrictive measure is needed to gauge performance when high-precision searches are required or when it is suitable to present users with only a few options; i.e., the "short list". Here, P(@5), defined as precision computed over the top <u>five</u> ranks, is used; in other words R-precision where R = 5[121].

$$P(@\ 5) = (RELret\ /\ Rknown\ |\ ranks = 1,2,..,5) \tag{9-4}$$

9.2.1.2 Results

The overall results (N=29 queries) are presented in Appendix B (Selected Precision Results) and summarized in Table 9-1. The summary includes both precision measures, P(@r) and P(@5), and three role-based cases where the relevance set, s-qrels, were composed of: *authorities (Auth)*; *hubs (Hub)*; and *either hubs and/or authorities (A/H)*. The P(@5) values are presented without bounds as the upper bound is, here, 1.0. It is important to note that the s-qrel pertaining

---

[120] Therefore, for each query and corresponding snowball sample, r = Rknown.

[121] If Rknown <5 Equation 10-4 is adjusted so that P(@5) is replace by P(@Rknown). This did not occur in retrieval runs using the current evaluation query set.

to any particular query (and for each precision measure) is adjusted to reflect computed role. For example, when computing precision, *Auth* s-qrels are reduced to members having positive authority scores. Similarly, *Hub* would filter the s-qrels to those having positive hub score. The last class in the table, *A/H*, treats as relevant, any actor with either a positive hub or authority score. User feedback suggests that *A/H* best reflects how the system should be assessed; essentially on it's effectiveness in finding authorities or hubs that may provide referrals.

While R-precision computed using the snowball sample is a rather "harsh" test of the system, the results are nonetheless encouraging. Taking the broadest relevance case, *A/H*, the mean R-precision is 37% across all test queries (Column 3); that is, *Expert Locator* and the snowball sample overlap by more than 1/3 based on the R-ranked retrievals. The P(@r) lower bound and upper bound values in the last two columns provide an interval that contains the "true" P(@r); the width of this interval reflects the uncertainty as to whether the system was finding additional relevant experts (novelty) not embedded in the snowball or retrieving non-relevant others. This cannot be resolved without further assessment. Moreover the main point here is that the system is performing well even at the lower bound and may be performing much better.

	mean P(@5)	mean P(@r)	P(@r) LB	P(@r) UB
Auth	0.641	0.313	0.313	0.842
Hub	0.552	0.444	0.444	0.706
A|H	0.793	0.371	0.371	0.943

Table 9-1: Summary Results

Mean P(@5), Column (2), is potentially most revealing as it measures the likelihood that any person ranked in the top five is relevant.   Informal discussions with users reflected the need to have high precision over the top five ranks with a strong probability of finding "experts"; i.e., authorities. Essentially, users wanted a highly accurate "short list". This would provide them with a reasonable first selection or could be used to identify alters; possibly in the same organization as the system user. Taking the most general case, *A/H*, short-list precision is quite high, 79%. Nearly four out of five actors is either an authority or hub on average. Interestingly, the system has a "preference" for ranking authorities over hubs (0.642 > 0.552) over the top 5 ranks.

9.2.1.3   *Implications for Signaling and Activity Theory*

Costly signaling theory (CST) posits that if signaling is costly and the Handicap Principle holds signaling is cost-prohibitive (or less likely at least) for actors with less skill and novices. Therefore, if CST holds, the expectation is that the probability that a signaler is an acknowledged expert is higher than the probability that signaler is a broker providing referrals primarily. In this case there is some evidence that CST holds here as from Table 9-2, P(@5) results show authorities (experts) are more likely to be in the top ranks than brokers (hubs). Recalling that the snowball sample establishes an organizational consensus as to who is an expert or broker, the results here suggest that signaling evidence is a reasonable predictor of expertise level given multiple types of experts.

Role	Mean P(@5)
Authority	0.642
Hub	0.552

Table 9-2: Precision as a Function of Role

Implicit in CST is the notion that signalers with a desired trait are more consistent in signaling that trait. For example, a top-level researcher is more likely to communicate skill level through a series of published papers than a researcher with less skill (other factors being equal). The sequential cost to signal is lower for the more proficient researcher. This suggests that in the current experiments across the top ranks, true experts (authorities) should be more consistent in signaling their expertise than brokers (hubs). To address signaling consistency, queries are blocked according to the two role cases: Authorities-only and Hubs-only. Then, the coefficient of variation, CV[122], is computed across all queries and for each role.   As shown in Figure 9-2, below, the system ranks authorities higher than hubs over the top 5 ranks and is more consistent in doing so (i.e., lower CV, 0.314 < 0.385).   As such, the results suggest that signal quality varies across signaler types; experts signal more effectively and consistently than brokers.

---

[122] CV is the standard deviation divided by the mean; used here as a rough measure of variation.
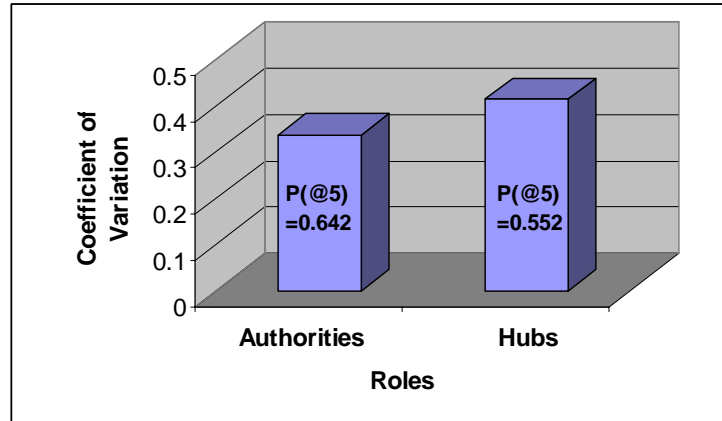
Figure 9-2:   Precision Variability as a Function of Role (using P(@5))

9.2.2   Experiment 2: The Effect of Activity Spaces on System Performance

Expert Locator is scalable; it is possible to adjust (add or remove) evidence used without changing the underlying scoring model.   The model uses a simple (weighted) linear model to aggregate evidence across activity spaces so that it is a straightforward process to add or remove activity spaces or to change weights of importance. However, it is not clear that adding or removing evidence will necessarily improve retrieval effectiveness; especially when using a large number of activity spaces.

Adding or removing activity spaces (sources) from the enterprise model is not an arbitrary process. Adjusting evidentiary sources used may bias the kinds of (expertise signaling) behaviors used; shifting emphasis towards either formal or informal work areas. In a relatively static work environment work directed through traditional management structure, and formal work spaces may merit significant emphasis. However, in dynamic organizations where expertise self-organizes around rapidly changing mission areas, informal work spaces like ListServs and community spaces may be more important for reflecting actual expertise. Therefore source selection, here in the form of enterprise activity spaces, is critical in determining what kinds of activities will be covered.

9.2.2.1 Approach

Using the advanced user's interface (in the research version), *Expert Locator* was run on all queries using various combinations of activity spaces. For example, the system was set up to run just the ListServ space. In that case only Listserv data were processed by the expertise model, and only the expert ratings based on ListServs were used to rank experts. Other spaces were treated as null spaces, and results fusion defaulted to simply using results on ListServs. Similarly, the system may be set up to run on only the Projects or Personal spaces. A more interesting question is how system performance varies when multiple activity space combinations are used. Regardless of configuration, retrieval performance is evaluated using the full query snowball sample since it is formed by survey-generated peer ratings that are not influenced by which sources of evidence are considered by the system.

Before examining the actual results, it is useful to describe the results table format. As noted, there are currently three activity spaces, which yield $\sum_{\forall r} C(n,r)$ (=7) combinations[123]; three single space variants, three cases involving two activity spaces, and one case where all three activity spaces are used. Table 9-3 provides the sensitivity analysis run using R-precision as the performance measure. The table is "stacked" with 5 layers (due to its size); each layer having the same format. An abbreviated query name is listed across the column heading and the seven activity space combinations are listed in column (1). (The reader is referred to Figure 8-4 for full query names.) The computed R-precision value for each query and activity space combination is entered into the appropriate cell and the highest scoring activity space combination is shaded gray. For example, the first query is *Air Traffic Control*, abbreviated as *ATC*. Note that R-precision is 0.14 for Lists (ListServs) and 0.03 for Pers (Personal) spaces. These are the two lowest. The highest R-precision was for the combination of Pers and Projs (Projects), R-precision = 0.24; and it is shaded. The table can be quickly scanned by looking for *Query* → at the start of the next block of 6 queries. The table provides a reasonable basis for assessing the robustness of the system to changes in evidentiary sources (activity spaces) across all (n=29) queries. The same format is used for the P(@5) measure and those results are found in Table 9-

---

[123] Where, C(n, r) is the combinatorics operator; and the summation is across all r space combinations (i.e. r=1,2,3)

4.   Note that the grand average across all queries for each activity space combination is provided in the last entry in Table 9-3 and Table 9-4.

It is worth noting here that a sensitivity analysis of this type is costly. Essentially, the system must be run and scored for each source combination across all queries. As described, above, this involves an extensive compilation in which query hits are compared to snowball generated relevance sets for each of (n=29) queries, across all seven combinations of sources and for both performance measures.   There are 2*7*29=406 separate analyses needed to cover this results space.

Query-->	ATC	Bayes Nets	BioComp	Biometrics	BrainMaping	Chem War
Spaces	P(@r)	P(@r)	P(@r)	P(@r)	P(@r)	P(@r)
Lists	0.14	0.26	0.05	0.29	0.06	0.20
Orgs	0.03	0.42	**0.40**	0.33	0.65	0.20
Projs	0.21	0.37	0.30	0.08	0.59	0.00
Lists/Orgs	0.17	0.53	**0.40**	**0.42**	0.47	0.30
Lists/Projs	0.21	0.47	0.30	0.29	0.47	0.20
Orgs/Projs	**0.24**	0.42	0.35	0.21	0.59	0.30
Lists/Orgs/Projs	0.21	**0.63**	0.30	**0.42**	**0.71**	**0.40**
Query-->	CAS	Geo-Map	GPS	Grid	HLS	HCI
Spaces	P(@r)	P(@r)	P(@r)	P(@r)	P(@r)	P(@r)
Lists	0.11	0.03	0.00	0.45	**0.31**	0.20
Orgs	0.31	0.21	**0.25**	0.35	0.22	0.35
Projs	0.13	0.00	0.00	0.10	0.04	0.22
Lists/Orgs	0.26	0.03	**0.25**	**0.60**	0.24	0.42
Lists/Projs	0.24	**0.24**	0.00	0.40	0.22	0.32
Orgs/Projs	0.30	0.21	**0.25**	0.30	0.22	0.42
Lists/Orgs/Projs	**0.33**	0.16	**0.25**	**0.60**	0.27	**0.45**
Query-->	IR	InsiderT	J2EE	Logistics	Nano	Net Prot
Spaces	P(@r)	P(@r)	P(@r)	P(@r)	P(@r)	P(@r)
Lists	0.34	0.48	0.35	0.15	0.36	0.07
Orgs	0.36	0.61	0.21	0.04	0.27	0.20
Projs	0.39	0.22	0.06	0.04	0.45	0.14
Lists/Orgs	0.37	**0.65**	**0.44**	0.11	0.45	0.23
Lists/Projs	**0.47**	0.39	0.31	0.19	0.50	0.18
Orgs/Projs	0.42	0.43	0.18	0.11	0.55	0.23
Lists/Orgs/Projs	0.46	0.61	0.41	**0.26**	**0.73**	**0.27**
Query-->	OR	Robotics	Sat Coms	SemWeb	SigProc	Sim&Mod
Spaces	P(@r)	P(@r)	P(@r)	P(@r)	P(@r)	P(@r)
Lists	0.20	0.23	0.00	0.32	0.07	0.13
Orgs	0.33	0.23	0.07	0.18	0.25	**0.27**
Projs	0.03	0.31	0.00	0.27	0.36	0.11
Lists/Orgs	0.30	0.46	0.10	0.32	0.18	0.33
Lists/Projs	0.17	**0.62**	0.05	0.32	0.23	0.21
Orgs/Projs	0.33	0.46	**0.15**	0.45	0.36	**0.21**
Lists/Orgs/Projs	**0.37**	**0.62**	**0.15**	**0.50**	**0.39**	**0.27**
Query-->	SNA	S/W-Eng	Speech	VegFor	Wearable	Grand Ave
Spaces	P(@r)	P(@r)	P(@r)	P(@r)	P(@r)	P(@r)
Lists	0.18	0.15	0.62	0.30	0.17	0.21
Orgs	0.24	0.08	0.65	0.50	0.17	0.28
Projs	0.47	0.14	0.58	0.00	0.17	0.19
Lists/Orgs	**0.59**	0.18	0.62	0.60	0.39	0.33
Lists/Projs	0.35	0.20	**0.73**	0.30	0.35	0.29
Orgs/Projs	0.41	0.23	0.65	0.50	0.39	0.32
Lists/Orgs/Projs	0.41	**0.27**	**0.73**	**0.60**	**0.43**	**0.39**

Table 9-3:   Sensitivity Analysis using P(@r)

Query-->	ATC	Bayes Nets	BioComp	Biometrics	BrainMaping	Chem War
Spaces	P(@5)	P(@5)	P(@5)	P(@5)	P(@5)	P(@5)
Lists	0.0	0.6	0.2	0.4	0.2	0.4
Orgs	0.0	0.6	0.6	0.6	0.6	0.4
Projs	0.4	0.6	**0.8**	0.2	**0.8**	0.0
Lists/Orgs	0.2	**1**	0.4	**1**	0.4	0.4
Lists/Projs	0.2	0.8	0.4	0.6	0.2	0.4
Orgs/Projs	**0.8**	0.8	0.8	0.4	**0.8**	0.4
Lists/Orgs/Projs	0.6	0.8	**0.8**	0.8	**0.8**	**0.6**
Query-->	CAS	Geo-Map	GPS	Grid	HLS	HCI
Spaces	P(@5)	P(@5)	P(@5)	P(@5)	P(@5)	P(@5)
Lists	0.6	0.2	0	0.6	1	0.6
Orgs	0.8	**0.8**	**0.6**	0.6	0.2	0.8
Projs	0.8	0	0	0.4	0.2	0.2
Lists/Orgs	0.8	0.2	**0.6**	0.6	**0.8**	0.8
Lists/Projs	0.8	**0.8**	0	0.6	0.2	0.8
Orgs/Projs	0.6	**0.8**	**0.6**	0.6	0.2	**1**
Lists/Orgs/Projs	**1**	**0.8**	**0.6**	0.8	0.8	0.8
Query-->	IR	InsiderT	J2EE	Logistics	Nano	Net Prot
Spaces	P(@5)	P(@5)	P(@5)	P(@5)	P(@5)	P(@5)
Lists	0.8	0.8	1	0.2	0.4	0.4
Orgs	0.8	**1**	0.2	0	0.4	0.8
Projs	0.6	0.4	0.2	0	**1**	0.6
Lists/Orgs	0.8	**1**	**0.8**	0.2	0.6	0.4
Lists/Projs	**1**	**1**	0.6	0	0.6	0.6
Orgs/Projs	0.8	0.8	0.4	0	**1**	0.8
Lists/Orgs/Projs	**1**	**1**	**0.8**	**0.8**	0.8	**1**
Query-->	OR	Robotics	Sat Coms	SemWeb	SigProc	Sim&Mod
Spaces	P(@5)	P(@5)	P(@5)	P(@5)	P(@5)	P(@5)
Lists	0.4	0.6	0	0.6	0.2	0.6
Orgs	1	0.6	**0.2**	0.6	0.8	0.6
Projs	0.2	0.4	0	0.4	0.4	0
Lists/Orgs	0.4	0.6	0	**0.8**	0.6	0.6
Lists/Projs	0	0.4	0	0.6	0.4	0.4
Orgs/Projs	0.4	0.6	**0.2**	**0.8**	0.8	**0.8**
Lists/Orgs/Projs	**0.6**	**0.8**	**0.2**	**0.8**	**1**	**0.8**
Query-->	SNA	S/W-Eng	Speech	VegFor	Wearable	Grand Ave.
Spaces	P(@5)	P(@5)	P(@5)	P(@5)	P(@5)	P(@5)
Lists	0.6	0.8	**1**	0.6	0.8	0.50
Orgs	0.6	0	**1**	0.8	0.4	0.57
Projs	0.6	0.6	**1**	0	0.6	0.39
Lists/Orgs	0.6	0.6	**1**	0.6	0.6	0.60
Lists/Projs	**0.8**	0.8	**1**	0.6	0.8	0.53
Orgs/Projs	**0.8**	**1**	**1**	**0.8**	0.8	0.68
Lists/Orgs/Projs	**0.8**	**1**	**1**	0.6	**1**	**0.80**

Table 9-4:   Sensitivity Analysis using P(@5)

9.2.2.2   Results

This section examines system performance with regard to the number of activity spaces used. For example, are 2-space configurations better performing than 1-space designs? Is prototype performance based on the 3-space design superior to 1-space or 2-space versions of the system? From Tables 9-3 and 9-4, above, mean P(@r) and mean P(@5)are computed for each of the three system configurations: 1-space, 2-space, or 3-space. For example, the average P(@r) for 1-space configurations is computed as the mean of the means; that is, the average of the average P(@r) across all queries and for the three cases: Lists, Projects, and Pers. Then, the average P(@r) = .227 = (.21 +.28 + .19)/3. As shown in Table 9-5, as the number of activity spaces used increases, retrieval performance improves for both measures. This monotonic behavior is desirable since it demonstrates the efficacy of adding activity spaces (at least for the limited activity spaces used in this experiment). The 3-space case shows reasonably high precision performance for both measures.   In fact, these results suggest the system can use fewer spaces if short high-precision lists are needed; however this is from a precision perspective and does reflect the impact of missing retrieval items due to missing spaces. This is especially true when a key person is omitted.

# of Spaces	P(@r)	P(@5)	P(@5) - P(@r)
One	0.227	0.487	0.26
Two	0.313	0.602	0.289
Three	0.39	0.8	0.41

Table 9-5:   Change in Performance with Numbers of Activity Spaces Used

The difference between P(@5) and P(@r), increases as the number of activity spaces increases, Column (4). There may be several factors contributing to this. First, this suggests that top five ranks will benefit more, precision-wise, from additional (relevant) activity spaces than lower ranks will. Essentially, outside the top five ranked positions, as R increases, the probability increases of retrieving nonrelevant actors. In addition, since snowball sampling does not guarantee coverage of the relevance population; missing relevance judgments or unknowns (*UNKret*) may also degrade precision performance since they are treated as "misses" and occur with higher frequency with increasing rank.

9.2.2.3 Implications for Signaling and Activity Theory

The premise here is that consistent with CST and Activity Theory experts are more likely to signal expertise in multiple relevant work domains than those with less expertise or non-experts. This is largely an extension of the simple sender-receiver asymmetric signaling model discussed in Chapter 6. Here it is scaled-up to level of multiple senders-receivers (an audience) across multiple work settings (activity spaces). Underlying this is the notion that experts central to an expertise network build trust and reputation through social interaction (costly signaling) across multiple forums. Reputation building is typically cost-prohibitive for novices and others with less expertise.   Based on results in Table 9-5, there is evidence that the probability of detecting experts increases with the number of relevant activity spaces. *That is, signaling across multiple forums is a predictor of expertise*.

Signaling is situated and the premise is that in some work contexts (Activity Spaces) costly signaling holds more reliably than in others. If this is so, there may be variation as to the extent that signaling in one AS is a better predictor of expertise than another AS. This is supported by the P(@5) results shown in Table 9-4; where from the single AS results, P(@5 is higher for Personal AS signaling than for either ListServs or Projects. This is summarized in Figure 9-3, below, where Personal space (Pers) has the highest precision scores. Figure 9-3 also includes performance from combined spaces and shows that the highest precision 2-space is the combination of the Personal and Project spaces. Interestingly, the Pers/Projs 2-space is the combination of the best and worst 1-space results. This suggests that even though one AS may be more effective for detecting expertise than another there may be redundancy or overlap in terms of experts found across two combined AS. As such, there is evidence that *signal cost varies with work context* and that the combination of signaling evidence from multiple contexts is likely sub-linear (i.e., experts are not typically unique to a single AS); however, this is very preliminary and further research is suggested.
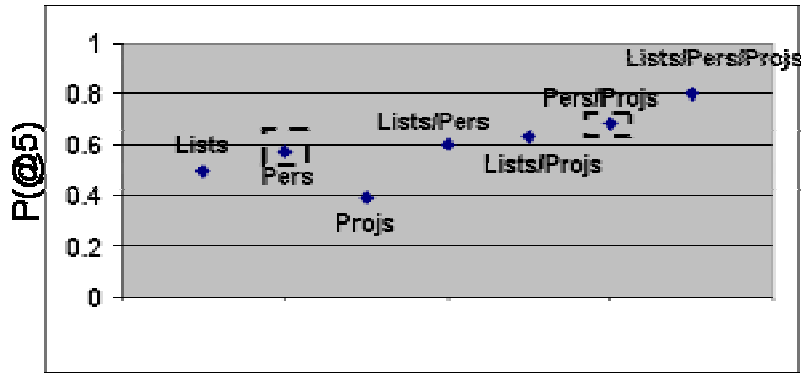
Figure 9-3:   Selecting the Best 1-space and 2-space Results

Activity spaces vary in their contribution to overall precision; signaling evidence in one AS may be a better predictor of expertise than another. This suggests AS weighting as a way to improve results fusion. However, preliminary investigation comparing uniform weighting to binary entropy weighting (see Chapter 7) showed, on average, little performance variation (<1%) between the two. While this by no means covers the spectrum of possible weighting schemes, a complete exploration of this result would require a larger set of queries and activity spaces in order to assess underlying factors. Interestingly, observations on a more limited test set indicate that binary entropy weighting and CombMNZ scaling largely serve to *shuffle* the composite list ranking but not influence precision scores from the original uniform weighting. Essentially, the number of relevant retrieved above cutoff, r, is nearly the same for both weighting schemes even though the rank order is often different. Again, a thorough investigation of AS weighting schemes is called for; and this is proposed for future work. In particular, further analysis should provide more insight as to the relationship between AS weights and signal cost.

9.2.3   Experiment 3: Does Precision Vary Across Queries?

From a user perspective performance variation may be evidenced in *missing experts* that lead to a loss of confidence in system coverage (similar to errors in known item searching), while in other cases, skewed rankings may evoke user concern, "why is Joe ranked higher than Mary?" In this section, performance variation across queries is explored. However, the very nature of the experiments, especially the problematic nature of obtaining relevance assessments, suggests that the nature of query variability cannot be sorted out fully. Part of the issue stems from the

treatment of *unknowns*; retrieval items that did not match known relevant or non-relevant. Unknowns cannot be resolved without additional (costly) relevance assessments; therefore, the assessment here remains conditioned on the assumption that unknowns are non-relevant.

To align this analysis with the standard system usage, only the *A/H* mode (Authorities or Hubs) is evaluated. Recalling the snowball sample based survey, the *A/H* mode views brokers (hubs) and experts (authorities) as relevant to the query, and this tracks user's view of relevance in terms of finding "experts" or referrals. The other two evaluation modes, *Authorities*-only and *Hubs*-only take a narrower view of the system and are not evaluated here.

9.2.3.1   Approach

Notionally, a core-periphery view is taken on query performance. Core queries have little variation and may be treated as a group; around this core are queries that have relatively low or high performance. The focus here is on identifying the core and periphery queries, isolating peripheral queries with low or high precision values, and then examining selected characteristics. This analysis is not designed to be complete in terms of exhaustively testing a wide range of performance-affecting variables; instead it is an initial investigation into sources of query variability. The following steps are taken.

1. Assess performance variability across queries for each precision measure: P(@r) and P(@5).

2. For a selected precision measure, identify "interesting" queries in the context of the overall evaluation set; that is, "low" and "high" performing queries.

3. Identify selected query characteristics.

Precision scores, P(@r) and P(@5), are computed across all queries for the *A/H* mode as shown in Figure 9-4; data values are provided in Table 9-6.   The Coefficient of Variation (CV) is computed across the query set for each precision measure. From inspection, there is considerable variation in P(@r) scores (CV = 0.40) compared to P(@5) scores (CV=0.21). As shown, P(@5) scores are limited largely to the range 0.6 to 1.0 with most values at P(@5)= 0.8; while P(@r) have a higher variance.   This suggests focusing on P(@r) only; however, there is dependency between P(@r) and P(@5); P(@r) uses information from the first five ranks just as P(@5) does.

186

However, Figure 9-4 does suggest that the dependency is weak, (r=0.2349, p-value=0.2201). This reinforces that precision over the first five ranks is not a strong predictor of precision over the first R ranks; again, where R is the number of known relevant. As such, since P(@r) has greater variation and can be treated separately from P(@5), the focus is on P(@r); less can be learned from analysis of P(@5).
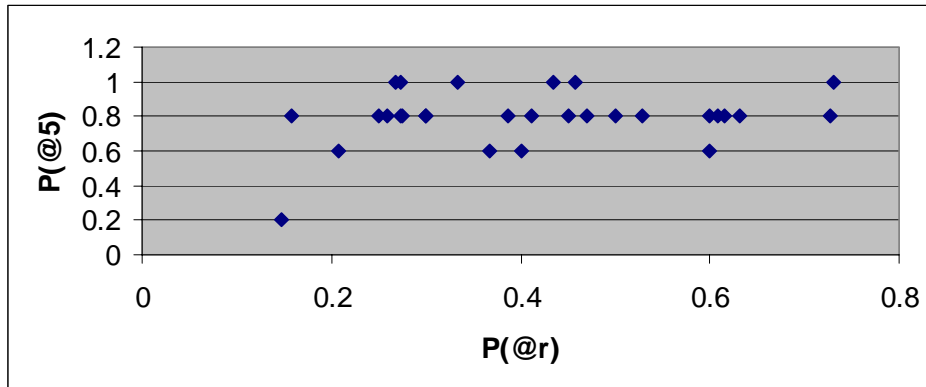

Figure 9-4: Co-variation in Precision Scores

Overall Performance (N=29)	A|H P(@r)	A|H P(@5)
Query		
Air Traffic Control	0.207	0.6
Bayesian Networks	0.632	0.8
Biocomputing	0.300	0.8
Biometrics	0.300	0.8
Brain Mapping	0.529	0.8
Chemical Warfare	0.400	0.6
Complex Adaptive Systems	0.333	1
Geospatial Mapping	0.158	0.8
Global Position System	0.250	0.8
Grid Computing	0.600	0.8
Homeland Security	0.275	0.8
Human Computer Interaction	0.450	0.8
Information Retrieval	0.458	1
Insider Threat	0.609	0.8
J2EE	0.412	0.8
Logistics	0.259	0.8
Nanotechnology	0.727	0.8
Network Protocols	0.273	1
Operationsl Research	0.367	0.6
Robotics	0.615	0.8
Satellite Communication	0.146	0.2
Semantic Web	0.500	0.8
Signal Processing	0.386	0.8
Simulation and Modeling	0.273	0.8
Social Network Analysis	0.471	0.8
Software Engineering	0.268	1
Speech Recognition	0.731	1
Vegetation Forensics	0.600	0.6
Wearable Computing	0.435	1

Table 9-6: Overall Precision for Both Measures

Box plots, Tukey (1977), are used to characterize the distribution of precision scores, using simple statistics. Box plots are exploratory data analysis tools used to discern patterns in scores and to identify outliers. In effect, Box plots are used to identify "core" and "periphery" queries. The Box plot in Figure 9-5 summarizes P(@r) across all (n=29) queries using five values. The left edge of the box is the 25th percentile, the line inside is the median, and the right edge is the 75th percentile. The two end lines reflect the minimum and maximum values in the data. The box represents the middle 50% of query scores (i.e., the core); therefore, queries on the periphery and falling outside the box contribute most to the variance in precision scores. Note, here, there

are six queries that have "low" precision scores, 1,8,9,16,21,26, and eight queries that have "high" scores, 2,5,10,14,17,20,27,28. The query labels are given in Table 9-7.
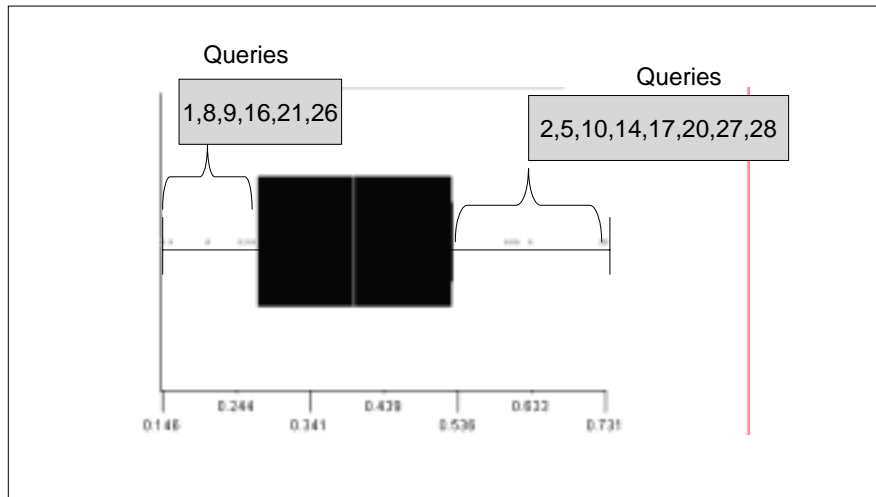

Figure 9-5: Box Plot for P(@r) Scores

Table 9-7: Potential "Outlier" Queries Contributing to Variability

Low and high precision queries are typed in terms of *specificity* and organizational *diffusion*. Specificity, computed here as inverse document frequency, provides a measure of topic distribution across the underlying information space (i.e., artifact evidence). This includes formal publications, ListServ postings, and various other documents. Expertise diffusion provides another view on the query and it relates to diversity; specifically the distribution of retrieved experts across organizational units (here, Divisions). Diffusion is measured here as

entropy; the higher the entropy in terms of the distribution of experts across organizational units, the more diffused expertise is across the organization.

The low and high performing queries are characterized by specificity and diffusion and plotted in Figure 9-6.   High performing queries are represented by a rectangle. <u>There is a clear pattern here where high performing queries have higher specificity and lower diffusion</u>. In other words higher precision is associated with queries that have relatively narrow query terms and relevant actors that are concentrated into fewer divisions.   The query *Geospatial Mapping* presents a mixed case as it has lower precision, high specificity, and high diffusion. In this case it is possible (although not conclusive) that the snowball sample was less effective in covering the relevant population and therefore did not reflect actual expertise diffusion. If so, the system exhibits wider "reach" than the two-wave snowball sample and therefore identified a relatively large number of *UNKret* [124]; resulting in lowers precision. Therefore, while the query *Geospatial Mapping* has high specificity it has lower precision and this may be do to the mismatch between the snowball sample and the system's ability to find experts in wide-ranging settings. Other explanations are possible here and further investigation, in part supported by extending the snowball sample seems warranted. On the other hand, the *Insider Threat* query had high precision, high specificity, and low diffusion and this is consistent with the narrowness of the topic and the concentration of work into only a few organizations.

---

[124] Note that in Appendix B, *Geospatial Mapping* had the highest percentage $UNKret$ (nearly 82%).
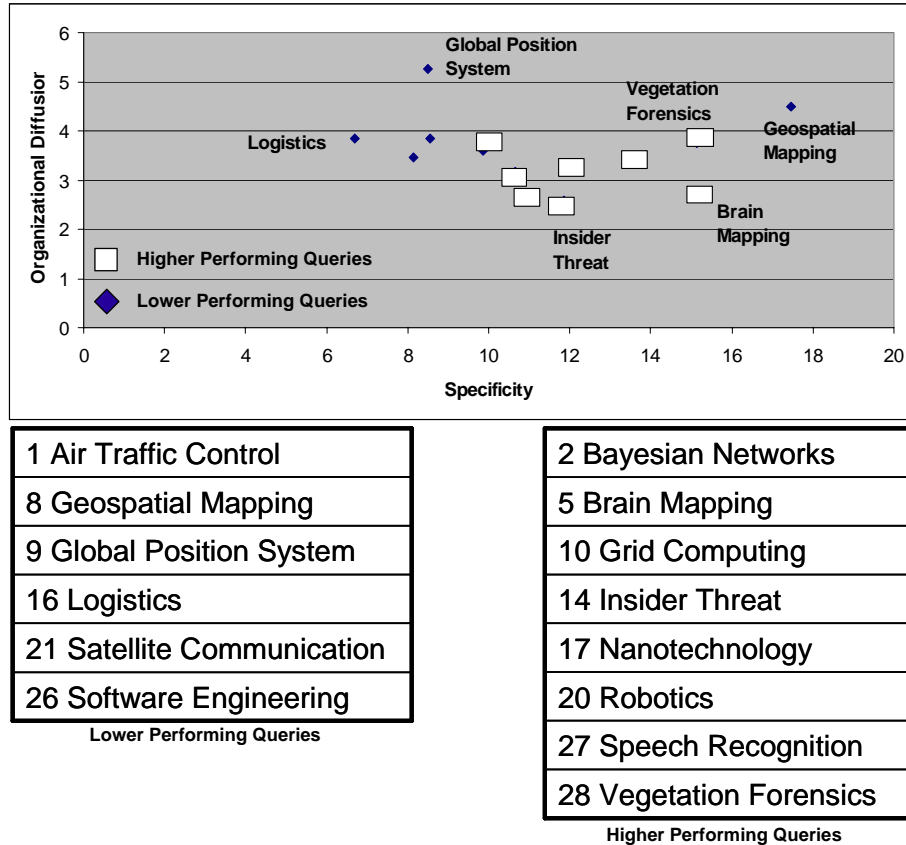
Figure 9-6: Low- and High-performing Queries, Categorized by Domain Characteristics

In summary, high precision queries are associated with expertise areas that are highly situated within the organization, involving a core group embedded into a relatively small number of organizations. Interestingly, these groups are also geographically concentrated; however that was not the focus of this preliminary analysis. Conversely, lower performing queries point to areas where experts are more widely dispersed across the organization and where the topic tends to be broad. For example, the query *logistics* is fairly general and is used across a range of resource and technology contexts. As a result, actors associated with its usage are relatively widely dispersed. Of course this raises context issues resolved by either providing a more specific query, for example, *logistic models*, or by providing social post-filters to increase the probability of locating true experts based on work context and organizational ties. Overall, typing according to specificity and organizational diffusion provides additional insight into sources of variability.

Low-precision queries are associated with ambiguous (low-specificity) signals that are not well correlated with the expertise trait; they are inherently unreliable. Of course, the application of signaling theory must address why some signals are reliable and others are not. As defined in Chapter 3 and 6, a signal consists of a basic theme and context. Receiver must assign reliability based on message content and the context in which the message is embedded. For example, assume sender signals expertise in *Global Positioning Systems (GPS).* Signal evidence maps to a particular activity space and receiver assesses reliability based on activity space evidence. If receiver is not knowledgeable in the expertise domain, receiver may be "deceived". For example, receiver may not be able to distinguish between the signal *Global Positioning Systems (GPS)* and the signal *Global Positioning Systems (GPS) used in smart weapons*; the latter likely signals expertise in a distinctly different domain. Of course, given the signaling model defined in Chapter 6, the expertise model is a proxy for receiver so that the *model* resolves signal reliability to the extent that the model can disambiguate signals.

9.3   Alternative Evaluation Measures

Expert Locator performance is based on *precision* without regard for whether retrieved experts are already known to a user. This can confound operational assessments where high-precision results may be largely redundant with user's a priori knowledge of who is an expert. In that case, precision, may <u>not</u> inform system utility where usefulness is based on the amount of "new" information provided. The potential disparity between accuracy and "information gain" suggests extending the current evaluation to address retrieval *novelty* as a measure of how much user is informed of experts not known prior to retrieval. Here, novelty, adapted from Korfhage (1977), is *the proportion of relevant retrieved experts that were previously unknown to the user*.

9.3.1   Background

Novelty detection has been widely addressed in information retrieval research to include recent work in TREC, Soboroff and Harman (2005), where the goal was to investigate methods to locate relevant, non-redundant information within an ordered document set. System accuracy

consisted of two aspects; accuracy in detecting relevant sentences (similar to passage retrieval) and the detection of new information (novelty). Performance was assessed using precision and recall (as combined in the *F*-measure). As reflected in the TREC studies, novelty detection is problematic; there is the intrinsic problem of judging relevance exacerbated further by the need to contrast current information with information already processed or with some reference *state*. This has implications for the *Expert Locator* evaluation in that novelty assessments must incorporate knowledge of who a particular user already knows is an expert.

Along those lines, Chen and Wu, (2006) used <u>knowledge of a user </u>to assess the novelty of knowledge discovery in the form of association rules. In effect the larger the semantic distance between a rule antecedent and consequence, the more novel the rule. Semantic distance was based on background documents associated with a particular user. The notion of juxtaposing system output to user's knowledge as a basis for discerning "new" information is conceptually consistent with the approach taken by Fujii and Ishikawa[125] (2000), who, in a document retrieval setting, measured the utility of one system by comparing it to another. They used a log ratio of detection probabilities, from two systems, to determine to what extent one system was producing novel results when compared to the second. This approach can be adapted to *Expert Locator* assessments where it is easy to show that novelty, as defined by Korfhage, can be cast as the ratio of two precision measures, P_a and P_n, as represented in the following:

$$N(s\,|\,q,u) = (RR'/T_{ret})/(RR/T_{ret}) = P_n/P_a \qquad (9\text{-}5)$$

where, $N(s\,|\,q,u)$ is the novelty of system, *s*, for query, *q*, and user *u*; $RR' =$ the number of relevant retrieved experts <u>unknown</u> to user; $RR =$ total relevant retrieved, and $Tret =$ total retrieved. Simplifying, novelty is the ration of P_n, the "novelty precision", and $P_a > 0$, the "accuracy precision" used in this thesis. However, novelty, as computed in Equation (9-5), departs from the current evaluation since while P_a is computed independent of any particular user; P_n is <u>dependent</u> on knowing user's "private knowledge" as to who is known to be an expert. This mandates a model of user's private knowledge describing *which experts a user knows*.

---

[125] citeseer.ist.psu.edu/593680.html Accessed on 19 December 2007.

User's private knowledge of which experts are relevant to a topic may be addressed in several ways; viewed here as a future extension to current experiments. First, and most direct, is a user-based evaluation in which selected users are directly involved in assessing relevance and "what's new". For example, using the snowball sample as a relevance baseline (as in current experiments), a user can identify which relevant retrieved are "new"; not previously known to user. In essence the snowball sample is used to assess relevance and the user judges novelty as a 2nd stage assessment. Then the novelty measure can be used to quantify the amount of new information retrieved for a particular. This can be repeated across user samples to generate an estimate of average novelty for the user population.

There is a special case in which novelty can be computed automatically; that is, the case where users are experts subsumed within a snowball sample. Essentially, the snowball sample will be used to profile user's private knowledge of known experts within a topic. This is illustrated through an example, below

9.3.2   Novelty Computed Using the Snowball Sample

By design, a snowball sample contains user's private knowledge of other experts. As such, the snowball supports not only the standard relevance judgments in current experiments, but also the identification of experts unknown to user. In the latter case, the snowball is used to determine the unknown relevant experts for a target user, RR', which is then used to compute P_n in Equation (9-5). Then, novelty is computed from Equation (9-5) as the ratio P_n / P_a. This is illustrated below using a sample query, the associated snowball sample, and simple graph overlap measurements.

An *Expert Locator* query is run on a particular topic and the retrieval list is folded into the snowball graph as represented in Figure 9-8. This is done by embedding an *Expert Locator* node in the snowball graph so that emanating arcs point to snowball members found by the system. Examining the *Expert Locator* node, $Tret = 15$, $RR = 12$ (i.e., there are (15-12 =) 3 nonrelevant nodes such as "Damianos, Laurie E" which is not in the snowball query relevant list, s-qrel.) With that, it is possible to compute retrieval novelty specific to any user in the snowball graph.

194

For example, for user ="Lehman, David", we can compute RR' as the intersection between user's ego graph and the system's retrieval graph so that $RR' = 12 - 3 = 9$. That is, of the 12 relevant retrieved experts 3 were known by Lehman[126]. With that $P_a = 12/15 = 0.80$ and $P_n = 9/15 = 0.60$. Therefore, novelty is $N(ExpertLocator \mid q, Lehman, Dave) = 0.60/0.80 = 0.75$. In another example, using "Gannon, Thomas", $N(ExpertLocator \mid q, Gannon, ThomasF) = 0.50$. *Since Gannon has greater awareness of relevant experts than Lehman, Gannon has lower novelty.* This is intuitive in that novelty is the inverse of awareness here; as awareness goes down, there is increased potential for novelty to go up.
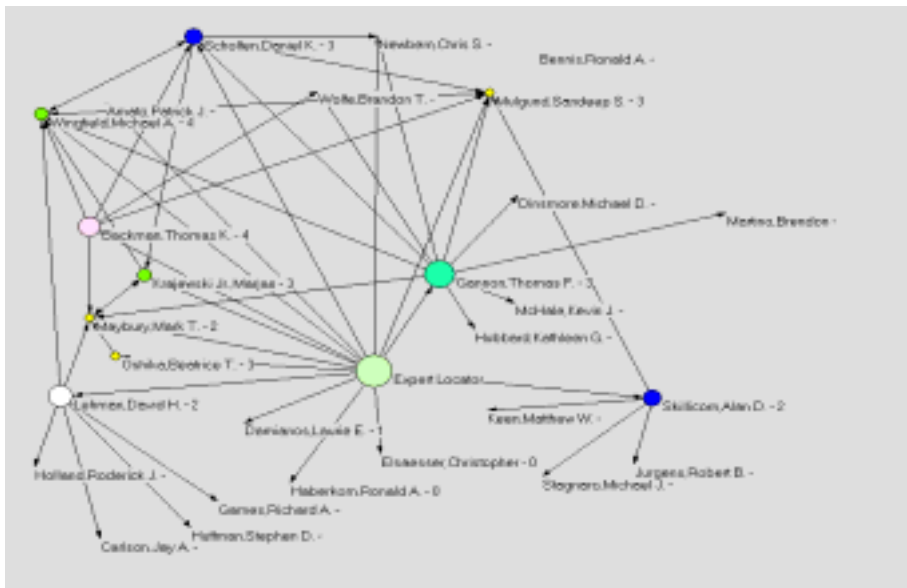

Figure 9-7 A Representative Snowball Graph with *Expert Locator* Incorporated

This approach can be extended to all test queries and related snowball samples so that it is possible in future work to extend current experiments to support automatic novelty assessments on all experts within the test query set.

9.3.3   *Novelty in the Context of the Current System Interface*

The notion of using novelty as a basis for assessing system performance suggests a reverse view; that is, to what extent can the system be engineered to enhance novelty without degrading

---

[126] This includes Lehman's self-awareness of his own expertise; otherwise there are two out links.

precision. If this is feasible then the system could be evaluated as to the extent it can predict which experts relevant to a query are known to a user. However, underlying this is the need to model user's "awareness" of others. While the goal here is not to redesign the system, it is useful to demonstrate the feasibility of develop a user awareness model that can be used rank relevant experts by their novelty. This will be explored briefly here.

The *Expert Locator* model is agnostic to user's private knowledge of actual experts. Therefore, while the thesis goal is to locate experts in disparate organizational settings the system is not "optimized" for high-novelty performance. To demonstrate the feasibility of building "awareness" into *Expert Locator* the following example is offered.

Using the prototype, a user, with employee ID #22882, generates a query: "Social Network Analysis". Assume that, internal to the system, two queries are run (although actual implementation will be more efficient than this suggests).

- The first search, as shown in the right retrieval graph in Figure 9-9, restricts retrieved experts to those having <u>organization or co-work ties to the target user, ID #22882</u>; this is done automatically using information from organization web pages and activity space membership. As indicated in the accompanying summary, 50% of the top 10, here colored red, were from the same division as user (and possibly had co-work relationships) and 50% (colors other than red) were known to have co-work ties only. This represents a group of experts that the user is likely to be aware of; so that for this retrieval novelty is by definition, zero.
- In the second instance, the user <u>relaxes</u> the personal network restriction so that retrieved experts are <u>not</u> required to have organizational or work ties (but may). This search may return experts outside user's personal network. Using the right-side graph to identify experts likely to be known by user, those that are <u>unlikely</u> to be known can be identified in the left-side graph. From that 20% of the nodes in the left graph are found to be from user's home organization and 40% have joint work with him. This leaves 40% that are not linked to the user's personal network; these are potentially experts user is unlikely to know. Using these simple statistics computed across the top ten ranked experts in the left graph, the estimated novelty score is 40%.

- The novelty score is essentially an estimate of the true novelty in that the profile may introduce error; that is there is likely some disparity between the user awareness model built automatically and user's actual private knowledge. All of which is to suggest the last experiment in which the user scores the novelty assessment. Here, a precision-like measure can be used to assess what percentage of retrieved experts judged "novel" is actually unknown to user.
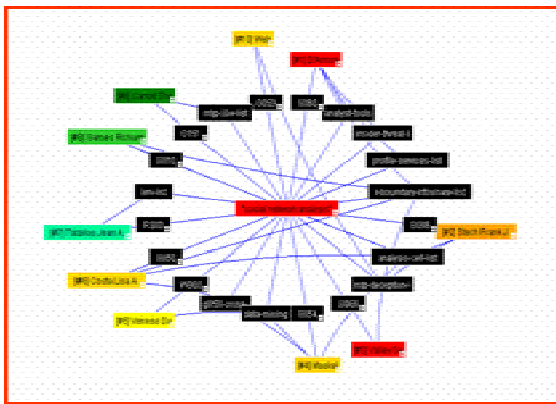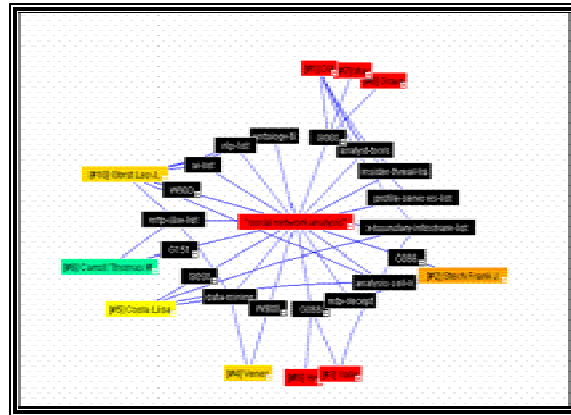
Figure 9-8:   Characterizing Retrieval Novelty

While, novelty measurement is central to this discussion; there is a balancing view that addresses *redundancy*.   This can be done from the perspective of system utility. That is, system utility is a function of the balance between novelty and redundancy; or similarly, between the amount of new and redundant information provided.   Here, novel information may increase user's awareness and knowledge of relevant others; while redundant information may be used to validate user's prior knowledge of *who knows what* and therefore build trust in system workings.

9.4   Structured Interviews

Interviews were conducted in two locations. The interviews were structured to determine the importance of expert finding in normal work and how people currently find and select experts; aligning with the notion that expertise location can be broken down into *Identification* and *Selection* phases Ackerman et al (1999). Identifying what kinds of information was used to find experts and what tools or methods were used was of special interest.

This survey[127] was time constrained and limited to a moderate (n=50) sample. While the survey may not be statistically significant, it does provide additional insight as to how expertise is shared within the organization.   A modified stratified sample was used where respondents were distributed across two dimensions: *AC level* and *years at MITRE*. The sampling frame consisted of 7 AC levels and 6 time bins covering MITRE employment: *less than 1 year, 1 to 2 years, 2 to 5 years, 5 to 10 years, and 10 to 20 years, and more than 20 years*.   The actual sampling distribution for each dimension is shown in Figure 9-10.
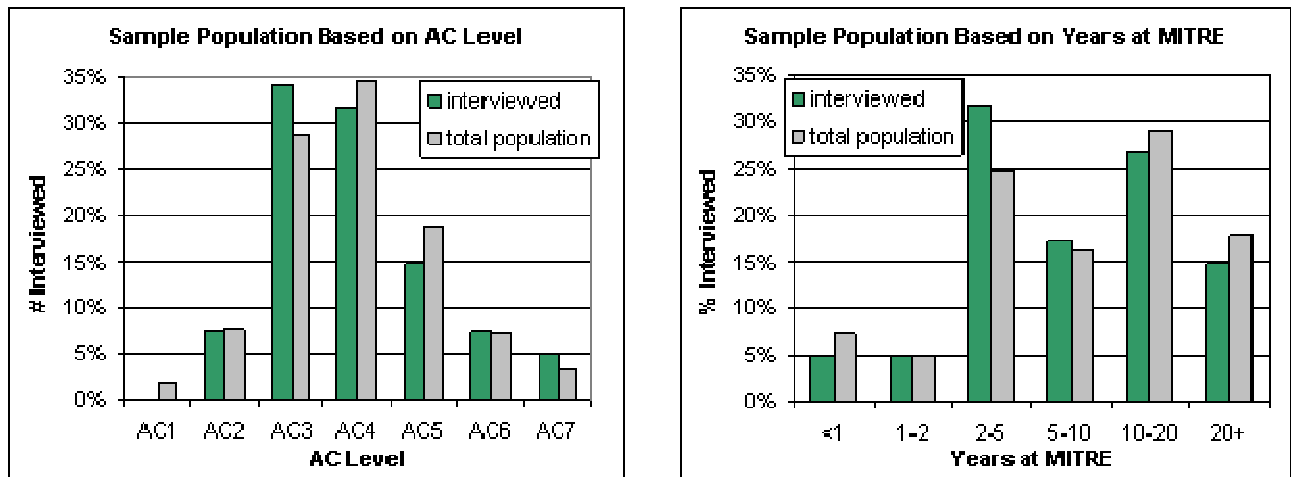

Figure 9-9: Sampling Distribution for AC Level and Years at MITRE

---

[127] Surveys were conducted by Raymond D'Amore (author) in MITRE's Washington facility and in the Bedford facility by a MITRE colleague.

9.4.1   Finding Experts

Without automated expert finding services, the most common methods cited for finding experts were asking colleagues for referrals (94%), and searching the Intranet for evidence (53%). When searching for experts in outside niche areas, referrals were less effective, and email broadcasts to selected work groups were more commonly used. For some users email queries were viewed as "risky" from the standpoint of exposing knowledge deficiencies to a potentially unknown internal community. This may restrict email use to queries that are viewed as "safely" distant from the user's actual or perceived area of expertise.

A number of more senior personnel viewed their work areas as "closed" domains in which there was certainty in terms of *who knows what* and not much need to search for expertise in other organizations. They tended to rely on their own personal networks. The corporate intranet also provided some support for finding experts through standard search services; for example, it is possible to query on-line collections for documents looking for authors relevant to an expertise area. However, this was not viewed as a very effective strategy. However, the Intranet did provide access to general expertise areas through the corporate InfoDesk or by browsing Technology Area Teams, formal technology groups set up to assess technology trends in industry and academia.

9.4.2   Selecting experts

For all people interviewed, selecting an expert from a list was not as simple as choosing the top name; expertise alone was not always considered to be sufficient. Most people would choose someone they knew and respected first. If they did not know anyone personally, they would base their decision on reputation followed by availability, physical proximity and employee AC level (capability). However, as one respondent noted: "*the higher the AC level, the less likely I am to contact that person; I won't ask stupid questions to higher ACs. They are better used for answering policy questions or questions on cross-disciplinary expertise.*" Some respondents considered availability and employee level to be negatively correlated; someone with more seniority was often regarded as being less accessible, available, or approachable than someone

with less seniority. Several respondents were more likely to contact someone within their own division first.

9.4.3   Using an Expert Finder

About three fourths of the participants surveyed had used an early prototype of *Expert Locator.* Independent of whether they had tried an expert finder before, more than three fourths surveyed would consider or would definitely use an expert finder capability. About one fourth would not consider using one or would use one very infrequently. Of the non-users, one employee was not enamored with finding others as much as needing an expert finder "*so that other people can find me.*" Some willing to use the system felt it would provide a way to identify people with like expertise outside their home organization or project base. A good example of this was team building; identifying staff to work on a particular project or part of an ad hoc study group.

9.4.4   Trusting an Expert Finder

Most people wanted to verify the performance of an expert finder by testing it on a topic with which they were familiar and examining the list of retrieved experts. People were not convinced that automatic techniques that mined resources available on the corporate Intranet would be adequate for finding and ranking experts. In particular, several employees were concerned that managers or project support staff may be incorrectly identified as experts on a particular topic just by association with people who reported to them or with other project members. Related to this, those that worked in restricted areas or primarily interfaced to outside organizations felt they would not be well represented in the system. Others were concerned that people would "spam" corporate work spaces (for example by frequently posting to a ListServ) in an attempt to increase their expertise rating. This suggests high precision is critical to building trust in an expert finder system and will dictate whether it is used frequently of not.   When asked what might make the system more "trustworthy" they responded that there is a need to "build in" some measure of reputation, potentially quantified by peer review, management awards, etc. Others wanted access to referrals much like they obtain from their personal networks. While automated

expert finding was viewed positively overall, amongst a subgroup there remains the need to support self-assessment to include providing skills descriptions and willingness to be contacted.

9.5   System Usage

Expert Locator is currently deployed as a test prototype for limited use. The system is not formally supported by MITRE except to provide server support to include maintenance and backup services. Further, the system is effectively "frozen" at this time; it remains a research testbed with limited user base, and not an operational system. *Expert Locator* users form a small group (about 10) of "first adopters" who are interested in the model-based approach but also are encouraged by the visualization interface that provides users with more flexible and intuitive ways to track expertise across disparate areas. As shown in Figure 9-11, below, there are approximately 36 queries per month on average (this computed over the first 11 months of 2007). Usage variance is in part due to lack of availability (e.g., June) and to some surge in use related to users' special application needs. While system usage and performance has not been formally assessed, it is expected there will be future assessments based on server log analysis, user interviews and other mechanisms.
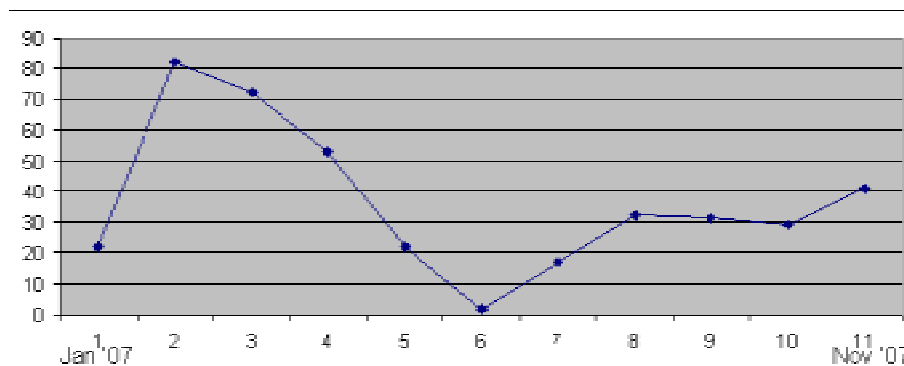

Figure 9-10: *Expert Locator* Usage (with minimal & uneven support)

Following the completion of this thesis, MITRE deployed an enterprise expert finder that was in part motivated by this research; however, the enterprise system is simpler in design in that it does not incorporate a formal expertise model nor does it exploit work context similar to the activity space constructs used in *Expert Locator*. The simplified design was motivated largely by the need to minimize system maintenance and to provide a basic initial service. For a given query, it

201

simply counts "hits" per person from the enterprise search engine, and ranks people based on total hits per query. Overall usage is shown in Figure 9-12, covering 11 months of 2007; there were on average roughly 2700 queries per month. While the purpose of this section is not to evaluate the initial enterprise capability (that is a future corporate activity) current usage reflects some level of organizational acceptance for expert finder services; and, potentially, provides an integration platform which may be used to integrate more advanced *Expert Locator* capabilities.
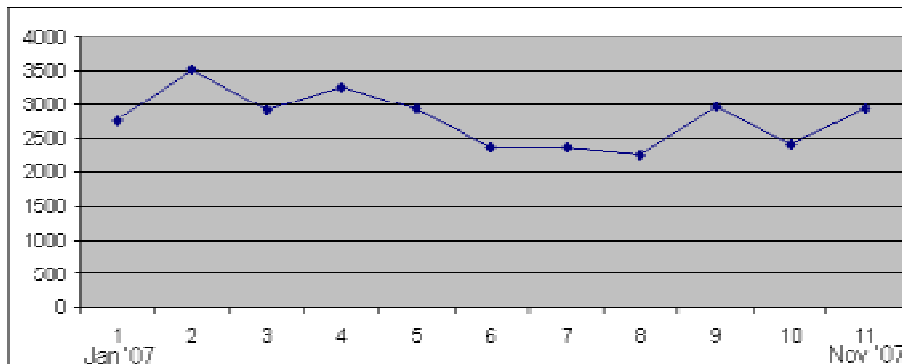

Figure 9-11: Usage Statistics for the Enterprise Expert Finder System

Finally, while the enterprise system is currently filling the need for a simple tool to find people and their associated documents; *Expert Locator* is causing some rethinking as to how these expert finder tools can be exploited on a broader basis. There is growing interest in using *Expert Locator* to support organizational network analysis for collaboration building, resource management, and in highly specialized applications such as insider threat detection. Some of these are discussed in Chapter 10.

9.6   Summary

Detecting experts within large heterogeneous environments can be problematic. As discussed, mission sensitivity, competition amongst knowledge workers and status risk makes expertise detection difficult; primarily through reduced expert signaling. This is particularly true for experts working in sensitive mission areas or in work spaces that cannot be instrumented for data collection due to technical or policy constraints. Nevertheless, the initial *Expert Locator* pilot has

demonstrated the potential for detecting expertise across a wide range of queries and based on evidence from only a few work spaces. In addition, system precision is reasonably well behaved. For example, a 2-point moving average of average precision as a function of rank is shown in Figure 9-13. Precision has an exponential decay; however interpolation yields roughly 77% precision at r = 10; 60% precision at r=15, and about 30% at r= 50. Most notable is high precision over the first 5 to 10 ranks; performance essential to user acceptance.
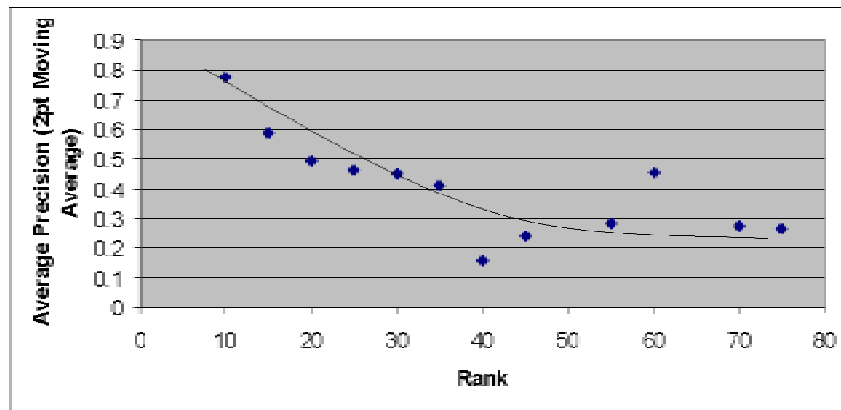

Figure 9-12: Average Precision vs. Rank (2-Point Moving Average—All Runs)

The system performance is monotonic across the number of activity spaces used. While there was little inter-configuration variation for example, all 1-spaces performed roughly the same, as did all 2-space combinations; additional activity spaces generally yielded higher precision. That is, 2-spaces generally outperformed 1-spaces and 3-spaces outperformed all other configurations.

Performance varied across queries. A preliminary assessment suggests that query performance may be influenced, in part, by query specificity as well as the diffusion of expertise across organizational spaces. Queries exhibiting low specificity and high diffusion performed more poorly than high specificity queries in which expertise was organizationally "localized".

Authorities, *true* experts, were more likely to be ranked highly (i.e., in the top 5) than hubs. This bias may be useful to the extent that users can rely on actual (hands-on) experts to be near the top of the retrieval list; similarly, it provides a rough way to organize retrieval results around "roles". For example, lower ranked actors are more likely to be "brokers" who have some knowledge of

the domain but are also likely to know who the true experts are. This can be coupled with organizational attributes such as home department to provide a preferred contact list; for example, tailoring selection to brokers or experts that are organizationally "close" to the user.

Since the completion of this research, the MII Intranet now includes a number of new potentially rich collaboration spaces that may provide evidence of expert signaling. For example, the MITRE Community Share initiative provides tools for work groups to establish collaboration spaces in which users can document their work, interact through discussion boards, send email and Instant Messaging, and make group work visible to others. Integrating Community Share into *Expert Locator* will likely provide additional new evidence of who knows what and improve retrieval performance.

While public spaces provide a range of work contexts to exploit, there is the need for methods that better insure privacy and control over information dissemination.   Some of this is currently being handled through (virtual) security enclaves for communities-of-interest and special projects. Bridging multi-security environments where information could percolate upwards from "low" to "high" access control levels is a critical need in certain multi-organizational environments. More generally there is concern that public information may be used to infer sensitive work and those involved. This *inference* problem is an obstacle to widespread dissemination of information and expertise sharing. In particular, the concern is that information "gaps" that protect identify may be filled by *distrusted others* who may be foraging on the periphery of sensitive areas. This follows some of the earlier research of Belkin, Oddy et al (1982) who viewed information seeking from the perspective of anomalous states of knowledge (ASK). Unless certain users can feel secure in sharing peripheral information, they are likely to avoid publicly signaling their knowledge.

10 Contributions and Future Work

This chapter outlines key contributions from the *Expert Locator* operational development and evaluation. This is followed by a brief discussion of potential future work.

10.1 Contributions

This research was carried out in a live operational environment; this necessitated an approach that balanced core research, with the need to align investigations with enterprise infrastructure services, corporate policy restrictions, and support for actual users. In addition, it motivated a more strategic view of expert finding; one that goes beyond narrow search issues, to instead cast expert finding as an element of organizational problem solving and work. This is reflected in the research contributions that follow.

10.1.1 A Survey of Expert Finder Systems and Models

The thesis includes a broad survey of the literature on expert finding that cross-cuts research investigations as well as commercial developments. The expert finder survey juxtaposes database-centric approaches, formed from user's self-assessments, to search and discovery paradigms that extend expert finding to environments that preclude formal registries or self-assessment as the primary *capture* mechanism. A number of systems and methods are described to include an extensive list of commercial enterprise and Web-based products. The survey culls out IR-based search methods aligned with the traditional query-answer paradigm, as well as specialized computational architectures to include agent-based, and peer-to-peer. The survey also points to exchange mechanisms emerging from the Semantic Web community that may be used to support cross-boundary (multi-organizational) expert finding.

From a more strategic view, expert finding is viewed as an element of organizational workflow. Here, expert finding is modeled as an adaptive process that incorporates various operations such

as profile/query generation, search, selection, and user feedback. While this model is agnostic as to a particular implementation strategy, it provides a framework for specifying *Expert Locator* functionality. Finally, the survey points to gaps in current research; in particular, the dearth of behaviorally-motivated expertise models. This supports investigation into the nature of expertise and expert behavior; providing the basis for a new class of expertise models.

10.1.2 A Survey Covering the Nature of Expertise and Signaling Theory

The IR literature on expert finding is decoupled from an extensive literature on the nature of expertise. As such, behavioral views of experts do not typically inform expert finding models grounded in information retrieval theory, for example. This motivates a survey of expert's behavior covering the cognitive science and decision-analytic communities from which a common element emerges: *experts signal their capabilities*. This motivates use of concepts from animal and human signaling theory as the basis for expertise modeling. A simplified signaling model that uses actor *activities* and work *context* is a cornerstone of the *Expert Locator* prototype.

10.1.3 An Activity Space Model

An underling premise in this thesis is that experts signal their qualifications through specific *activities* and *artifacts* within some organizational setting. As such, the central unit of analysis is the *activity space*; a sampling frame of sorts that binds expert signaling behavior to a particular work context. Activity spaces are grounded in Activity Theory which provides a rich conceptual space from which to address context and specific work elements such as actors, community, division of work, and mediated actions. The link between *actions* and *signals* within an activity space is a foundation of the *Expert Locator* model discussed in Chapters 7 and 8.

10.1.4 An Expertise Model Informed by Signaling Theory and Activity Theory

The *Expert Locator* model provides an extensible framework for evidence aggregation and expertise ranking. The model is informed by signaling theory in that signaling evidence (in the

form of artifacts and social interaction) forms the basis for rating expertise. The *Expert Locator* model provides an extensible framework for evidence aggregation and expertise ranking; it differs from most extant systems in that organizational context is represented directly using activity spaces tailored to reflect characteristics of a particular work setting. Architecturally, the system is viewed as a multi-agent decision model in which each activity space is associated with a decision agent. Decision agents synthesize signaling evidence from multiple activity spaces into actor rankings. The model can be extended to incorporate new activity spaces and model parameters can be adjusted at query time or fixed as system defaults. Most importantly, evidence types are explicit in the model so that final rankings reflect the relative importance of say, artifacts, relationships, or other factors consistent with default settings or user preferences.

10.1.5 An Operationally Deployed Expert Locator Prototype

Expert Locator is currently deployed in an operational environment. *Expert Locator* provides a free-text query interface similar to numerous Web retrieval systems but augments this with interactive visualization tools used to explore expertise networks. This visual interface coupled with supporting evidence and organizational context supports *selection* which is a function largely unaddressed in existing commercial systems and research prototypes. The system serves a dual purpose; it is an operational prototype as well as a research and evaluation testbed; as such the system is positioned to support new capabilities, and user-based evaluations not easily accommodated in "offline" evaluations such as TRECENT. Testbed utility has been demonstrated in that it now supports a number of new applications enabled by the *Expert Locator* core capabilities. This includes personal network management tools used to characterize key individuals in the context of their work relations.

10.1.6 An Operational Evaluation

Operational evaluations are relatively rare with regard to IR, in general, and expert finding more explicitly. The approach taken here is embedded within a live operational setting so that it exploits corporate infrastructure and actual experts. The evaluation is aligned with existing infrastructure and work practice; to include policy restrictions regarding the scope or specificity

of the actual testing. A novel snowball sampling scheme is used to generate a consensus-based relevance graph that provides a query-specific relevance set. This survey-based approach counters inefficiencies introduced by random sampling designs, and obviates the use of fixed or assigned expert panels used to assess relevance. In effect, actual experts and others with expertise awareness are used to assess the system.

Precision-based performance measures were used, in part, to reflect user's preference for short, high-precision retrieval lists. The evaluation was designed to assess system performance as a function of work coverage; and, not surprisingly, retrieval performance increases with the number of activity spaces used. However, system performance varies across queries; and sensitivity testing reveals that the highest performing queries are characterized by high specificity topic terms and low organizational *diffusion*. That is, the highest rated experts are located within organizational niches and not scattered across disparate organizations.

10.1.7 Evaluation from a Signaling Theory Perspective

From an expertise signaling perspective, there is support that costly signaling theory holds across the various experiments and that signaling evidence is a predictor of expertise. In addition, there is evidence that signal quality varies across signaler types; in particular, there is support for the assertion that experts signal more effectively and consistently than brokers who provide referrals primarily. The signaler-receiver asymmetric game model developed in this thesis applies to signaling within multiple contexts in that precision increases with the number of relevant activity spaces. That is, signaling across multiple forums is a predictor of expertise; multi-forum signaling is *costly* and aligns with the *handicap* principle.

There is precision variation across queries that suggest signal reliability varies across topical domains; broad-based domains, which may have multiple meanings, and are potentially, organizationally dispersed, are likely to have lower signal reliability. The opposite argument can be made for high-precision queries that are associated with high-specificity signaling domains and are generally associated with organizational niches. This is finding is preliminary; further investigation is warranted.

Signaling is situated and the premise is that in some work contexts (activity spaces) costly signaling holds more reliably than in others. If this is so, there may be variation as to the extent that one AS is more useful for finding experts than another. This is supported by the precision results that indicate significant performance differences across signal activity spaces. In addition, there is evidence that while one AS may be more effective detecting expertise than another, for combinations of AS the increase in precision is likely sub-linear. This is due to overlapping membership; for example, the top 2 AS do not combine to produce the top 2-space precision results. Future research here might address questions such as what kinds of experts are likely to use a particular forum.

The notion that activity spaces vary in their contribution to overall precision, suggests AS weighting as a way to improve results fusion. However, a preliminary investigation comparing uniform weighting to binary entropy weighting (see Chapter 7) showed, on average, little performance variation (<1%) between the two. This is at odds somewhat with the notion that precision performance varies across 1-space AS and various AS combinations. However, these results are very preliminary; additional investigation is warranted.

10.2 Future Work

Future work divides between prototype enhancements and exploration of new research applications. A brief discussion of each area follows:

10.2.1 Prototype Enhancements

- **Extended Functionality**: As with many first-generation prototypes, operational use drives change requirements. Many proposed modifications address ease-of-use; while others suggest integration of whole new functionality. Most of these are outside the scope of this research in that they fall under the purview of configuration management groups tasked to maintain and adapt the system. In addition, a number of research extensions are deferred due to constraints in the current environment; for example, various low-level text analysis and indexing strategies, typically applied at the collection

level, are not easily addressed in the current architecture given corporate requirement to integrate with the corporate search engine API and the need to comply with organizational information access policy. Essentially, from the standpoint of text indexing the search engine is a *blackbox* so that methods, such as entity extraction, are restricted to retrieval post-processing. As such, in the immediate future, this precludes addressing a number of interesting problems such as actor-artifact attribution and, most importantly, expertise profiling.

Expertise is not persistent in the current system; it is made visible in response to a particular query. This is a potential limitation in that knowledge of experts gained during retrieval is effectively lost or at least not *shareable*. This could be addressed, in part, through expertise profiling which could be used to generate a persistent (adaptable) expertise signature, made to be shareable and managed. However, as noted above, this is precluded in the short term due to policy restrictions. Alternatively, knowledge of experts could be captured and exploited using relevance feedback and expertise tagging. Relevance feedback is largely unexplored in expert finder systems. The notion here is to exploit user feedback, in the form of query-specific expertise ratings, as the basis for creating a knowledge directory of sorts. While the overall feedback model is not specified here, a bottom-up approach would address the efficacy of rating lower-level evidence in order to enhance overall retrieval or selection. This may consist of *typed* feedback where user assessments of artifacts (traditional relevance feedback) may provide a basis for selectively weighting artifact evidence, while feedback on candidate experts and their ties to others (social relevance feedback) may be used to modify the weight assigned to activity space contexts they inhabit.

Relevance feedback results may be stored for future use potentially. While there are research issues related to the rating schemes used, the more pressing issues may be privacy related. As discussed earlier, user ratings may be inaccurate and misused which would likely lead to loss of trust and nonuse of the system. The research focus here needs to address *privacy-enhanced relevance feedback*.

- **Activity Space Modeling**: *Expert Locator* model uses relatively simple activity space models sufficient to capture artifact and selected social evidence. There are bases, however, where activity space models can more fully account for actor behavior at the individual or group level. For example, a potential enhancement to the implementation of the ListServ activity space model is to incorporate a thread segmentation scheme that reliably breaks connected postings into subsequences based on shifts in post content or poster's organizational locations.   This may better ensure threads are homogeneous as to content and social composition; which may enhance precision at the retrieval end.

10.2.2   Evaluation Enhancements

The evaluation has underpinnings in three main areas: test query generation, specifying query-relevant sets, and context coverage. Viewed as a three-dimensional cube each experiment *edge* may be extended, notionally, so as to scale-up the overall evaluation.

- **Test Query Set Generation**: A key evaluation issue centers on system robustness with regard to query variability. The current evaluation was based on roughly 30 queries sampled from various work domains; however, while there is persistence in many work domains, new areas or important variations in current topics emerge continuously. As such, a future focus is to expand the evaluation query set to include additional queries from core areas as well as queries from sparsely populated domains.

- **Expanded Snowball Sampling**: The snowball sampling scheme coupled with the HITS algorithm was used to develop a typed relevance set; where relevant experts, viewed as nodes within a consensus graph, were categorized as to role: *authorities* or *brokers*. The proposal here is to expand the snowball graph size for each query beyond the two hops generated in the original survey in order to develop higher-coverage qrels. A practical starting point is to run the snowball until some "simple" stopping condition is reached (e.g., until a policy forced termination.). However, more interesting is the notion of developing an information gain like measure that is used within a snowball convergence strategy. Essentially, when the amount of new "information" (i.e., additional experts)

exceeds acquisition cost, *stop*. Clearly the existence of "local minima" conditions must be addressed so as to avoid early convergence; this suggests a dynamic snowball convergence scheme that is able to "adjust" the survey protocol in order to reduce the possibility of missing experts.

- **Context Expansion**: The current expertise model computes query-actor similarity based on a linear combination of evidence from multiple activity spaces (AS). As discussed in Chapter 9, retrieval performance increases with the number of activity spaces; the three-space model outperforms the 2-space model, etc. However, an open question arises as to activity spaces selection; essentially, which combination of spaces ensures "good" performance; while this is not of particular importance in the current scheme, limited to three activity spaces, this may change as more activity spaces are added. This leads to a context selection problem; if N activity spaces are available which M out N spaces are "optimal" for a given query or for queries in general. Here the notion of "optimal" may take the form of *minimal cost per unit of retrieval precision*; however, based on usage models, a *satisficing* approach, tailored to reflect the trade-off between high precision *searches* and *selection* diversity may be required.

- **Other Performance Measures**: The current evaluation is largely precision/recall based; however other measures that incorporate user's a priori awareness of experts may be more effective in assessing system utility in an operational environment. Underlying this is the notion that subject matter experts engaged in the relevant domain may know many of the key experts in an enterprise; as such a high-precision search may not necessarily provide *new* information. This motivates future experiments that measure retrieval *novelty*; supporting ideas are discussed in Chapter 9.3.

- **User Feedback**: User input played a key part in this research. Early on users provided insights into expert finder requirements; they made visible which corporate services currently provide expert finding support and how services were being used, they provided queries to support preliminary testing; and most importantly, they were supportive of the evaluation survey and actively participated in follow-on prototype use. However, there is

significant opportunity to involve users more fully as the prototype goes through iterative enhancements and, potentially, future assessment.

Long term success depends heavily on understanding how users adjudicate evidence of expertise in order to determine who is an actual expert; although the system provides evidence in the form of social context and artifacts relevant to a query; users may use other cues *as to who has relevant expertise*. In addition, there are areas where users vary in their level of system trust; for example, some activity spaces have higher reliability in terms of the signaling evidence found; that is, it is known to be a better indicator of true expertise. Here users are unwittingly building their own *costly signaling model* which raises a more general issue of the need for strategies that effectively combine user and system evidence weightings.

Privacy issues may weigh heavily on whether the system will be used long term. For example, in a few cases users will work around leaving online "footprints" as to the nature of their work—in some cases this is unavoidable since it relates to concerns over work sensitivity, in other cases it reflects broader privacy concerns regarding creating centralized stores of expert's work behavior. As such, user discussions will be critical to developing strategies for managing private versus public knowledge of user's expertise; this is especially important where expertise ratings from peers are captured.

10.2.3 Future Applications

Expert Locator provides users with "new" ways to locate expertise without having to resort solely to broadcast email, face-to-face interactions, or other traditional means for getting advice. However, the system has potentially wider use in terms of adding expertise as context to traditional social and organizational network analysis. For example, organizational network analysis used to identify workflow in some operational domain may be viewed from the perspective of expertise embedded in specific tasking. Here, *Expertise Locator* may be used to generate an expertise overlay used to identify key actors or skill areas associated with specific work activities and products. Integrating expertise detection with traditional organizational network analysis may provide whole new ways to address resource allocation especially in

identifying scarce skill areas across projects. Several examples are presented below; first, *Expertise Locator* is used directly to generate an expertise network which is analyzed using standard social network analysis methods. In this case the expertise network is specialized community of practice (COP) in which the network is homogeneous with regard to some skill area common to members. Network analysis and visualization is performed with standard social network tools[128] residing outside of the *Expertise Locator* toolset. The second case is focused on finding expertise embedded within a heterogeneous COP. Finally, the third case focuses on the analysis of technical exchange meetings where expertise detection can be used to identify the juxtaposition of key experts, assigned presenters, and the links to other participants.

10.2.4   Expertise Networks as a Specialized Community of Practice

Tracking extant or emerging communities of practice has significant importance across the enterprise. Community formation may signal the emergence of a new technology area or collaboration on a particular problem or customer base, Maybury, D'Amore, and House (2001). Community detection may support resource allocation; for example, the juxtaposition of internal research funding with emerging research areas may support research planning and be used to set funding priorities. *Expert Locator* has been used to identify community structure and evolution in which community members possess a common (expertise) trait. For example, there is interest in identifying emergent work in malicious insider detection; i.e. identifying the connections between all staff members working the insider threat[129] problem.

Insider threat is an increasingly important area motivated by both sponsor and organizational requirements to address enterprise security. While it is ostensibly a physical and information security issue, the problem may be viewed from several vantage points to include information seeking behavioral models, topic analysis, and vulnerability analysis as it relates to insider behaviors such as surveillance (scanning the work environment) and social engineering (e.g.,

---

[128] Various social network tools are viable here; for example, Pajek, http://vlado.fmf.uni-lj.si/pub/networks/pajek/. Accessed on December 15 2005

[129] Insider Threat is "a rogue employee or malicious hacker who has gained access to internal networks by obtaining legitimate credentials." www.intrusic.com/WhatThreat.htm Accessed on November 15 2005

targeting key persons to co-opt). Given the interdisciplinary nature of the problem, it is difficult to track researchers working this problem; i.e., to ferret out the actual community of practice.

For the insider threat domain, *Expert Locator* is set up to run a number of queries (this may vary from a simple phrase; e.g., "malicious code insertion", to a set of related queries). The system retrieves artifacts and social context (i.e., actors, activities, and organizations) relevant to the topic and organizes the overall results as a retrieval graph. System output is directed to standard graph analysis packages for analysis of the overall graph. The retrieval graph can be viewed as an expertise community network and take several forms. For this example, a single mode graph is generated where nodes represent activity spaces (e.g., projects) and edges reflect co-membership. Figure 7-10 depicts the Insider Threat community graph where nodes have several shapes; *circle*= Project, *rectangle*=ListServ, and *diamond*= Department. Edge thickness represents the level of overlap between activity spaces (i.e., the number of people jointly involved with the connecting activities)[130].

There are a number of simple metrics that may be used to characterize communities, such as centrality and density. While these types of metrics are used fairly regularly in social network analysis to characterize network structure or significant nodes, of greater interest here is community evolution or "state". More specifically, a long term focus is on developing a community *maturity model* that casts community evolution along a business continuum of sorts. Stated simply, at one end there are emergent communities that are often loosely coupled, may focus on wide-ranging issues not central to current mission focus, and generally lack formal support (e.g., related project work). However, over time, community may become more cohesive and its focus may converge with corporate mission, become coupled more directly with formal organization structure, and align or influence strategic direction. This evolution from fragmented community to one that is synchronized with the formal organization may occur over several phases.

However, in the absence of a formal maturity model, community "state" or evolution is characterized through an assessment of various subgraphs. For example, in Figure 7-10, the "1"

---

[130] In addition, node size is scaled according to nodal centrality computed using Pajek.

signifies a sub-graph in which there is self-organization around two key ListServs: INFOSEC and INSIDER-THREAT.   In addition, there are several internal research projects (as indicated by the "2" on the graph); one internal research project is focused on insider threat *user behavior data collection and analysis*. Other projects reflect the multi-disciplinary aspects of insider threat research; for example, *information management* is reflected by the "4", and *denial and deception*, by "3".   The graph components identified so far suggest an emerging community that reflects self-organization (around specific ListServs) as well as formal organization support represented by internal research projects.

The large connected network component is centered on the Insider Threat List Serve and G20 Division/Departments which has the corporate charter for leading work in the Insider Threat domain. Within this core, there are several organizations connected to the main activity spaces to include the G022, G020, and G021, all are information security organizations. Clearly, there are a number of formal organizations that are disjointed; not connected to the main graph component. Follow-up here indicates that these organizations are tracking insider threat work programs for various customers/sponsors but are not heavily involved in addressing the problem at this time.

The community periphery shows evidence of diffusion into new project areas that are not coupled into the core work. For example, the HDIS-List, "5", is an "island" separated from the main community subgraph. HDIS is focused on a particular sponsor problem area not integrated into the main community focus. This is interesting in the context of community maturity in that islands or isolates in the community graph can signal structural holes or discontinuities that may evidence restricted information sharing or poor work program coordination.
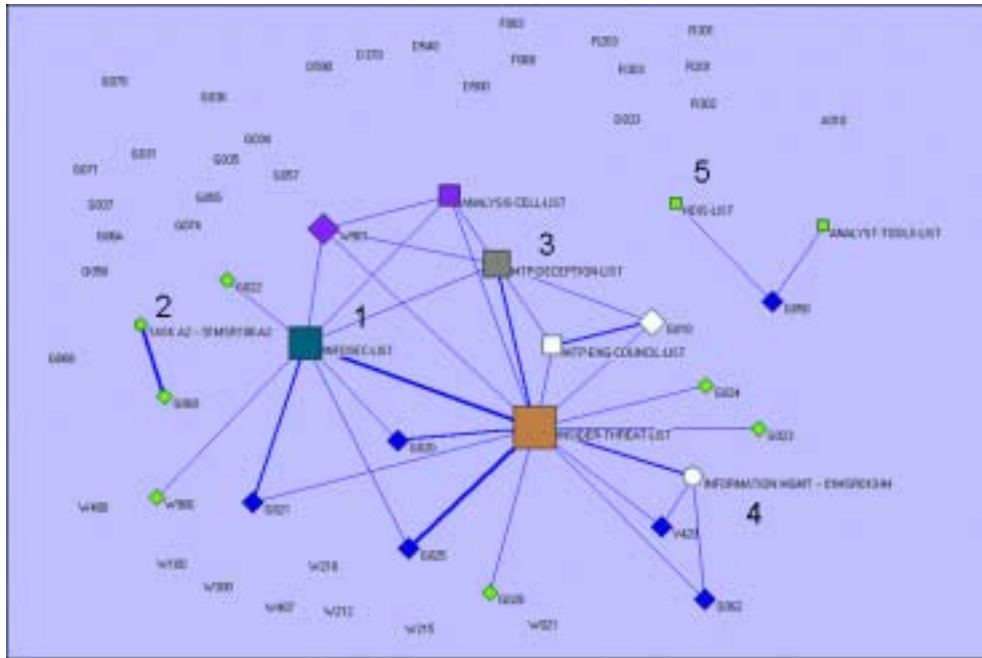
Figure 10-1: Detecting Embedded Expertise with a Community of Practice

10.2.5 Expertise Networks Embedded within Heterogeneous Communities

In this application, *Expert Locator* is used to detect actors with specific expertise that are embedded in a broader community graph. Detecting embedded expertise may be especially useful in resource allocation (or assignment) problems where the focus is on optimal use of scarce resources. A number of operational questions typically arise; for example: who should be assigned to certain tasks; how should scarce skills be distributed across program areas or geographic locations to maximize interaction with sponsors; and where are there critical skill gaps. An example follows.

Expertise detection tools can be used to automatically tag actors as to their skills while showing their network position within general program areas. Figure 7-11, illustrates a large, multi-program area covering wide ranging technologies, sponsors, and locations. A resource manager may need to know where certain skills are assigned; here the query is "data mining". As such, *Expertise Locator* was used to identify staff members with data mining expertise and to identify where they were assigned within the targeted program area. Note that data mining experts are represented by the large triangle nodes (with node size proportional to expertise level). In this

217

case, there are several experts distributed across only a few programs. In addition, several programs have at least two of the data mining experts assigned to tasks. This suggests that the few experts are clustered around only a few programs and that there may be data mining work not well supported in other programs in this area. At this point resource managers can determine if there is a need to reallocate these resources, acquire additional experts for use in other programs, or simply to increase communication across programs not currently exploiting this scarce resource. This may be done using Technical Exchange Meetings.
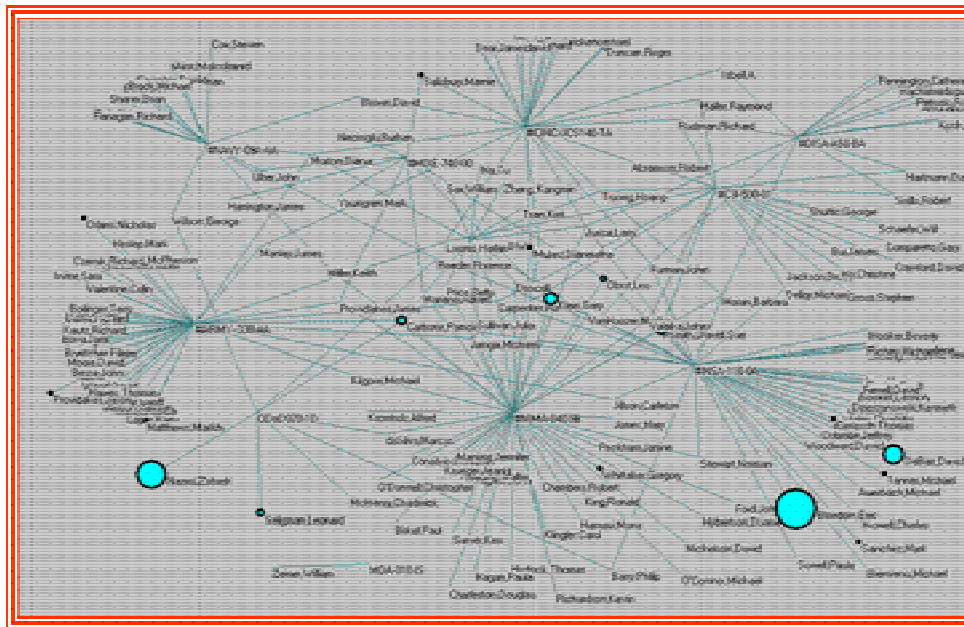
Figure 10-2: Embedded Expertise.

10.2.6   Technical Exchange Meetings

Many organizations use technical exchange meetings (TEMs) as a means to focus key issues, establish working groups, and form communities of practice. While TEMs are open to all interested individuals, they are organized much like formal conferences or workshops in that, individuals register to attend, special topics are culled out, and key note speakers and other presenters are identified. Identifying which papers or topics to address at a TEM can be problematic as organizers must identify who is doing what and from that, select papers that are central to the theme of the TEM. Here *Expert Locator* may be used to detect key researchers

within the TEM theme areas who may serve as TEM organizers or discussion group leaders; or in a narrower instance, be used to assess how nominated speakers relate to known experts. This later case, juxtaposing speakers to known experts, will be addressed through an actual application.

TEM organizers put together a TEM on malware (i.e., malicious software) using *Expert Locator* to baseline community membership. A call for papers went out to the enterprise and a committee selected papers relevant to each TEM sub-theme. In addition, an online registration process was used to sign up attendees. First, *Expertise Locator* was used to generate a "malware" expertise network which was compared to actual registration. Figure 7-12, left image, shows the network position of selected speakers (larger triangle nodes) within the retrieved network; clearly, selected speakers are on the network periphery. The overall graph shows a dense main component with several subgraphs with weak ties to the community core. There are a few disconnected digraphs and one of the speakers is an isolate not connected to any other network members. As such, organizers may now assess whether selected speakers are representative or not of core community work. There is also a bias to selecting speakers from the subgraph in the lower left side of the graph. This may or may not be aligned with the organizers intent.

The right side graph highlights highly ranked experts (i.e., node sizes reflect rank). Contrasting the left and right side graphs, most experts are not presenters and are situated in other areas of the community graph. As a rough first analysis, further questions may be addressed; for example, is the TEM exposing new non-central topics, focusing in on perceived shifts in interests, and are there emerging experts in the community as reflected in the selected speakers and papers.
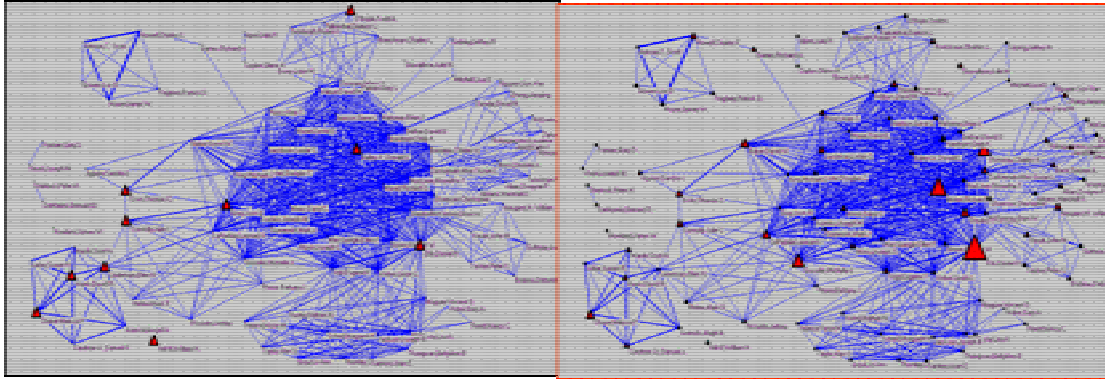
Figure 10-3: Partial Overlap between Selected Speakers (left graph) and Experts.

There are other uses of the *Expert Locator* output. For example, the retrieval graph could be used to predict attendance with the possibility that network position or certain actor attributes may be useful in predicting who will attend a particular TEM. The predictive model could be assessed using actual attendance. In addition, the difference between retrieved experts and attendees may be used to asses "missing experts". In particular, attendees not retrieved by the system could point to missing evidence associated with some activity space not currently integrated into the expertise model.

The current prototype is suggestive of a potentially new class of personal network management (PNM) tools (viewed here as a component of personal information management). These PNM tools can be used to glean organizational work, to identify key individuals, and to support team building and collaboration. They go beyond simply managing personal information but provide a type of social computing support used to assess network embeddedness and implications of work performance. For new employees this can be valuable in terms of identifying key people, projects, and informal groups. It can provide alternate paths to expertise that mitigate perceived risk in exposing knowledge gaps or lack of social ties to supervisors or peers. However, the current prototype has limited support for personal network management and considerable work remains to evolve the current tools towards providing more robust search, visualization support, and collaboration tools for building teams and supporting joint work.

The ability to locate key expertise across consortia or government agencies, for example, is of great importance in mission critical areas dependent on information sharing and access to subject matter experts not necessarily resident within a single organization. For example, while there is increasing focus on expert finding methodology; especially within TRECENT, there is emerging interest in making expert finding interoperable across disparate organizations. As noted in Chapter 2.5, the Semantic Web and related efforts are focused on infrastructure to support cross-boundary information exchange. Central here is the development of dictionaries and ontology that motivates future work in developing cross-boundary expert finders that bridge multiple organizations and environments.

The initial view of experts, Chapter 3, raises a dilemma. On one hand, research shows that experts fill key roles in organizations and possess critical skills and domain knowledge; while in other settings, experts perform poorly especially in certain prediction or estimation tasks. The notion that experts may under perform in certain settings has become the center piece of recent investigation into collective intelligence, prediction markets, and the *wisdom of crowds*. The popular view of this, as presented by Surowiecki (2004) and others, is that in selective cases *the many are smarter than the few*; even if the few are renowned experts and the many are a disparate group with widely varying knowledge. Essentially, crowds composed of diverse individuals that have private knowledge, act independently, operate in a decentralized manner, and have some collective basis for "deciding" may outperform even the best individual experts. While this perspective is backed up with supporting examples, it is also contrasted by cases where group think and other information inefficiencies dominate; then crowds may perform poorly. This area (and the debate surrounding it) is outside the scope of this thesis however, it raises a number of issues for future investigation. Essentially, there is a need for research at the intersection of collective intelligence and individual expertise. While research here may focus on the conditions where groups outperform individuals; from an expert finding perspective, the emphasis may be on finding collective knowledge diffused across a group or community and not centered within a single expert. While this is an area for future research, in the short term *Expert Locator* adheres to the notion that "crowds have their place, but experts live in niches, and for businesses that is where the real value lies.[131]"

---

[131] http://www.professorbainbridge.com/2006/week48/index.html Accessed on 11 September 2007.

11 References

[1]   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.

[2]   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).

[3]   Ackerman, M. S. and Halverson, Christine H. (2003) "Sharing Expertise: The Next Step for Knowledge Management." Wulf, Volker and Marlene Huysman (eds.) Social Capital and Information Technology. MIT Press, in press, 2003.

[4]   Ackerman, M. S. and Halverson, Christine H. (2004) "Organizational Memory: Processes, Boundary Objects, and Trajectories." Computer Supported Cooperative Work: The Journal of Collaborative Computing, April 2004, 13 (2), pp. 155-189.

[5]   Akerlof, G., (1970) The market for lemons: quality uncertainty and the market mechanism. Quarterly Journal of Economics 84:488–500, 1970.

[6]   Albrecht, J. W., and Ours, J.C. van (2006). Using employer hiring behavior to test the educational signaling hypothesis. Scandinavian Journal of Economics, 108(3), 361-372

[7]   Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Sheth, A., Arpinar, B., Ding, L., Kolari, P., Joshi, A., and Finin, T., (2006) Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection. In WWW 2006, Edinburgh, Scotland.

[8]   Allen, J. and R. van. de. Velden, (2005). "The Role of Self-Assessment in Measuring Skills". Paper for the Transition in Youth Workshop 8-10 September, AskMe Enterprise, http://www.askmecorp.com, accessed July 2004.

[9]   Aslam, J. A., and Montague, M., (2001) Models for metasearch, Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, p.276-284, September 2001, New Orleans, Louisiana, United States

[10]   Azzopardi, L., Balog, K., and de Rijke., M., (2004) Language modeling approaches for enterprise tasks. In Proceedings of the Fourteenth Text REtrieval Conference (TREC 2005), Voorhees, E. M., and Buckland, L. P. , editors. Gaithersburg, MD, November 2004.

[11]   Bacharach, M., and Gambetta, D. (1997). Trust in signs. In K. S. Cook (Ed.), Trust in Society. (pp. 148-184.). New York., Russell Sage Foundation

[12]   Backes-Gellner, U., and Arndt, W., (2006) "Entrepreneurial Signaling: Success Factor for Innovative Start-Ups" (March 2004). http://ssrn.com/abstract=383941 accessed on August 15, 2006.

[13]   Baeza-Yates, R., and Ribeiro-Neto, B., (1999) Modern Information Retrieval, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, 1999

[14]   Balog, K. and de Rijke, M.. Determining expert profiles (with an application to expert finding). In Proceedings IJCAI-2007, 2007.

[15]   Balog, K., Azzopardi, L., de Rijke, M., (2006). Formal models for expert finding in enterprise corpora. In Proceedings of the 29th annual international ACM SIGIR Conference on Research and Development in Information Retrieval. Pages 43-50.

[16]   Bao, S., Duan, H., Zhou, Q., Xiong, M., Cao, Y., Yu, Y., (2006) "Research on Expert Search at Enterprise Track of TREC 2006", in Proceedings of the 15th Text Retrieval Conference (TREC 2006)

[17]   Bartel, B. T., Cottrell, G. W., and Belew, R. K. (1994) Automatic combination of multiple ranked retrieval systems. In W. B. Croft and C. van Rijsbergen, editors, SIGIR'94, Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 173–181, Dublin, Ireland, July 1994. Springer-Verlag, London.

[18]   Bartel, B. T., Cottrell, G. W., and Belew, R. K. (1994a). Automatic combination of multiple ranked retrieval systems. In Proceedings of the 17th Annual Intl. ACM SIGIR Conf. on Research and Development in Information Retrieval.

[19]   Becerra-Fernandez, I. Facilitating the online search of experts at NASA using Expert Seeker People-Finder. In Reimer, U., editor, Proc. of the Third Int. Conf. on Practical Aspects of Knowledge Management PAKM 2000.

[20]   Becker, H. S. (1982) Art Worlds. Berkeley: University of California Press, 1982.

[21]   Becks, A., Reichling, T., and Wulf, V.: Supporting Collaborative Learning by Matching Human Actors, Proceedings of the 36th Annual Hawaii International Conference on System Sciences (2003)

[22]   Bedard, J., and Chi, M.T.H. (1993). Expertise in Auditing. Auditing: A Journal of Practice and Theory, 12, 21-45.

[23]   Belkin, N. J., Oddy, R. N., and Brooks, H. M., Ask for information retrieval: Part i: background and theory. Journal of Documentation, 38(3): 145-164, 1982.

[24]   Berg, S., Snowball Sampling, in Kotz, S. and Johnson, N. L. (Eds.) Encyclopaedia of Statistical Sciences Vol. 8.

[25]   Bergstrom, C. T., and Lachman, M. (1998) Signaling among relatives. III. Talk is cheap. Proceedings of the National Academy of Science USA 95: 5100-5105

[26]   Lachmann M., C. T. Bergstrom, and S. Számadó (2001). Cost and conflict in animal signals and human language *Proceedings of the National Academy of Sciences USA*. 98:13189-13194.

[27]   Bliege Bird, R., and Smith, E. A., Signaling theory, strategic interaction, and symbolic capital. Current Anthropology 46(2): 221-248

[28]   Bloch, F., Rao, V., and Desai, S., "Wedding Celebrations as Conspicuous Consumption : Signaling Social Status in Rural India," Université catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES) Discussion 1999022, Université catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES), 1999.

[29]   Brin, S., and Page, L., (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine, in Proceedings of the 7th International World Wide Web Conference Vol.7, 1998

[30]   Buckley, C., and Voorhees, E. M., (2004) Retrieval evaluation with incomplete information. In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2004), pages 25–32, Sheffield, UK,, July 2004. ACM Press.

[31]   Callan, J. P., Lu, Z. and Croft, W. B., (1995) Searching distributed collections with inference networks, in E. A. Fox, P. Ingwersen and R. Fidel, editors, 'Proceedings of the ACM SIGIR International Conference on Research and Development in Information Retrieval', ACM., 1995.

[32]   Cao, Y., Liu, J., Bao, S., Li, H., (2005) Research on expert search at enterprise track of TREC 2005. In Proceedings of the Fourteenth Text REtrieval Conference (TREC 2005), Voorhees, E. M., and Buckland, L. P. , editors. Gaithersburg, MD.

[33]   Chase, W. G. and Simon, H. A. (1973) Perception in Chess. Cognitive Psychology, 4, 55-81.

[34]   Chen, X. and Wu, Y. B. (2006). Personalized Knowledge Discovery: Mining Novel Association Rules from Text. Proceedings of SIAM SDM'06, April 20-22, 2006, Bethesda, MD, USA

[35]   Chi, M., Hutchinson, J., and Robin, A. F. (1989). How inferences about novel domain-related concepts can be constrained by structured knowledge. Merrill-Palmer Quarterly, 35, 27-62.

[36]   Chi, M., Glaser, R., and Farr, M. (eds.) (1988). The Nature of Expertise. Lawrence Erlbaum Associates, Publishers

[37]   Clark, H. H. (1992) Arenas of Language Use. University of Chicago Press, Chicago, IL

[38]   Clark, H. H., Schreuder, R., and Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. Journal of Verbal Learning and Verbal J. Morgan (Eds.), Syntax and semantics 3: Speech Behavior, 22, 245–258.

[39]   Clearwater, S., Huberman, B., and Hogg, T., (1992) "Cooperative Problem Solving" Computation: the Micro and the Macro View, B. A. Huberman editor, 33, World Scientific Publishing (1992).

[40]   Coibion, M., (1995) Surgical approach to subclinical breast lesions, Rev Med Brux. 16 (4):218-24 7481228, 1995.

[41]   Cole, M. and Engeström, Y., (1991) A cultural-historical approach to distributed cognition. In G. Salmon (Ed.) Distributed cognition, Cambridge: Cambridge University Press, 1991.

[42]   Converse, J. and Presser, S. (1986) Survey Questions: Handcrafting the Standardized Questionnaire, Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-001. Sage Publications, Beverly Hills, CA, pp 40-41.

[43]   Cormack, G. V., Palmer, C. R., Clarke, Charles, L. A., (1998) Efficient construction of large test collections, Proceedings of the 21st annual international ACM SIGIR Conference on Research and Development in Information Retrieval, August 24-28, 1998, Melbourne, Australia.

[44]   Craswell, N., de Vries, A. P., Soboroff, I., (2005) Overview of the TREC-2005 Enterprise Track, http://trec.nist.gov/pubs/trec14/papers/ENTERPRISE.OVERVIEW.pdf Accessed on 15 July 2007.

[45]   Craswell, N., Hawking, D. Hawking, Vercoustre, A. M., and Wilkins, P., (2001) P@noptic expert: Searching for experts not just for documents. In Ausweb, 2001.

[46]   Craswell, N., Zaragoza, H., Robertson, S., (2005) Microsoft Cambridge at TREC-14: Enterprise track. In Proceedings of the Fourteenth Text REtrieval Conference (TREC 2005), Voorhees, E. M., and Buckland, L. P. , editors. Gaithersburg, MD.

[47]   Croft, W. B., (2000) Combining approaches to information retrieval. In TREC 2 [1], pages 35-43.

[48]   Curley, S. P., Subramani, M. R. and Peddibhotla, N. B., (2002) How Do I Know That You Know? Factors Used to Infer Another's Expertise and to Communicate One's Own Expertise Eight Biennial Behavioral Decision Research in Management Conference, Chicago, Illinois, May 31-June 1, 2002.

[49]   D'Amore, R. (2004) Expertise Community Detection, Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval 498--499 , (2004)

[50]   D'Amore, R., (2005) Expertise Tracking, 2005 International Conference on Intelligence Analysis, https://analysis.mitre.org/proceedings_agenda.htm Accessed on 20 June 2006.

[51]   Damianos, L., Griffith, J., and Cuomo, D., (2006) Onomi: Social Bookmarking on a Corporate Intranet, WWW 2006 Tagging Workshop.

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

[53]   Davenport, T. H., and Prusak, L., (1998). Working Knowledge: How Organizations Manage What They Know. Cambridge, MA: Harvard Business School Press.

[54]   Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., Perrier, L., (2006) Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA 296: 1094-1102, 2006.

[55]   Davis, J., (2006) Improving CIA Analytic Performance: DI Analytic Priorities, The Sherman Kent Center for Intelligence Analysis Occasional Papers: Volume 1, Number 3, Sherman Kent Center.   https://www.cia.gov/csi/kent_csi/docs/v01n3p.htm Accessed on 17 July 2006

[56]   Dawes, R. M., and Corrigan, B. (1974). Linear models in decision making. Psychological Bulletin, 81, 95-106.

[57]   De Groot, A. D., and Gobet, F., (1996) Perception and memory in chess, Assen, The Netherlands: Van Gorcum.

[58]   De Long, (2002) Better Practices for Retaining Organizational Knowledge: Lessons from the LeadingEdge, Accenture Institute for Strategic Change, Research Report.

[59]   DeVaro, Jed, and Waldman, Michael, (2005) The Signaling Role of Promotions: Further Theory and Empirical Evidence, Working paper, Cornell University.

[60]   Dixon, N. M., (2000) Common knowledge: How companies thrive by sharing what they know, Boston, Mass: Harvard Business School Press.

[61]   Dreyfus, H.L. and Dreyfus, S.E., (1986) Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer. Free Press.

[62]   Du Boulay, B., and Ross, P., (1991) MSc Intelligent Tutoring Systems, School of Cognitive and Computing Sciences, University of Sussex.

[63]   Ebbesen, E., and Konecni, V., (1975) Decision making and information integration in the courts: The setting of bail. Journal of Personality and Social Psychology, 32, 1975.

[64]   Elsayed, T., and Oard, D. W., (2006) Modeling Identity in Archival Collections of Email: A Preliminary Study, 3rd Conference on Email and Anti-Spam (CEAS), pp. 95-103, Mountain View, July 2006.

[65]   Engeström,Y., Engeström, R. and Vähäaho, T., When the Center Doesn't Hold: The Importance of Knotworking. (1999) In S. Chaiklin, M. Hedegaard, and U. Jensen (eds.), Activity theory and social practice: Cultural-historical approaches. Aarhus, Denmark: Aarhus University Press.

[66]   Epple, D., Argote, L., and Murphy, K., (1996) An empirical investigation of the micro structure of knowledge acquisition and transfer through learning by doing, Operations Research, 44, 77-86, 1996.

[67]   Ericsson, K. A., and Polson, P. G., (1988) An experimental analysis of the mechanisms of a memory skill. Journal of Experimental Psychology: Learning, Memory and Cognition, 14, 305-316, 1988.

[68]   Ericsson, K. A., Krampe, R. T., and Tesche-Römer, C., (1993) The role of deliberate practice in the acquisition of expert performance. Psychological Review, 100, 363-406, 1993.

[69]   Finkel, E., Heath, C., and Dent, J., (2001). Expertise and the curse of knowledge: The communication problems of specialists. Working paper. University of North Carolina at Chapel Hill.

[70]   Fischer, G., (2000) Social Creativity, Symmetry of Ignorance and Meta-Design. Knowledge-Based Systems Journal (Special Issue on Creativity and Cognition), Elsevier Science B.V., Oxford, UK, 13(7-8), 527-537, 2000.

[71]   Fischer, G., and Ostwald, J., (2001) Knowledge Management — Problems, Promises, Realities, and Challenges, IEEE Intelligent Systems, January/February 2001, pp. 60-72.

[72]   Foddy, W. (1993) Constructing Questions for Interviews and Questionnaires: Theory and Practice in Social Research, Cambridge University Press, Cambridge, UK, pp 60-75.

[73]   Foner L. N., (1997) "Yenta: A Multi-Agent Referral-Based Matchmaking System", in Proceedings of the First International Conference on Autonomous Agents (Agents'97), Marina del Rey, CA..

[74]   Foss, J. E., Wright, W. R., and Coles, R. H. (1975). Testing the accuracy of field textures. Soil Science Society of America Proceedings, 39, 800-802.

[75]   Fox, E. A., and Shaw, J. A., (1994) Combination of Multiple Searches. Paper presented at the 2nd Annual Text Retrieval Conference (TREC-2), NIST, Gaithersburg, MD, 1994.

[76]   Fox, E. A., Koushik, M. P. Shaw, J., Modlin, R., and Rao, D. (1993)Combining evidence from multiple searches. In D. Harman, editor, The First Text Retrieval Conference (TREC-1), pages 319–328, Gaithersburg, MD, U SA, Mar. 1993. U.S. Government Printing Office, Washington D.C.

[77]   French, J., Powell, A., Callan, J., Viles, C., Emmitt, T., Prey, K., Mou, Y., (1999) Comparing the Performance of Database Selection Algorithms, in Proceedings of SIGIR '99, pages 238-245.

[78]   Friedkin, N., Horizons of Observability and Limits of Informal Control in Organizations. *Social Forces 62*, 1, (1983), 54-77.

[79]   Fu, Y., Yu, W., Li, Y., Liu, Y., Zhang, M., Ma, S., (2005) THUIR at TREC 2005: Enterprise track. In Proceedings of the Fourteenth Text REtrieval Conference (TREC 2005), Voorhees, E. M., and Buckland, L. P. , editors. Gaithersburg, MD.

[80]   Gaines, B. R., (1988) Positive feedback processes underlying the formation of expertise. IEEE Transactions on Systems, Man and Cybernetics SMC-18(6) 1016-1020, 1988.

[81]   Gambetta, D., and Hamill, H., (2005) Streetwise: How Taxi Drivers Establish Their Customers' Trustworthiness, New York: Russell Sage Foundation.

[82]   Gigerenzer, G., Todd, P., and the ABC Research Group, (1999) Simple Heuristics that Make Us Smart. New York, Oxford University Press.

[83]   Glaser, R. (1986). On the Nature of Expertise. In F. Klix and H. Hagendorf (Eds). Human memory and Cognitive Capabilities: Mechanisms and Performances (pp: 915-928). Elsevier Science Publishers.

[84]   Glaser, R., and Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser, and M. Farr (Eds.), The Nature of Expertise. Hillsdale, NJ: Erlbaum.

[85]   Goffman, E. (1959). The presentation of self in everyday life. Garden City: Doubleday-Anchor.

[86]   Goldberg, F., (1970) Toward expansion of Witherspoon: Capital scruples, jury bias and the use of psychological law. Harvard Civil Rights-Civil Library Law and Review. 5, 53-69, 1970.

[87]   Granovetter, M. (1973). The Strength of Weak Ties. American Journal of Sociology 78 (May): 1360-1380.

[88]   Guilford, T. C., and Dawkins, M. S., (1991) Receiver psychology and the evolution of animal signals. Animal Behavior. 42: 1-14, 1991.

[89]   Hall, B. G., (2006) The Health Care Quality Improvement Act of 1986 and Physician Peer Reviews: Success or Failure?,   http://www.semmelweis.org/articles/success%20or%20failure.pdf Accessed on September 11, 2006

[90] Hansen, M., (1999) The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. Administrative Science Quarterly, 44(1), 82-111, 1999.

[91] Harman, Donna, (1993) Overview of the first TREC Conference. Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Korfhage, Robert, Rasmussen, Edie, and Willett, Peter, eds., New York: ACM Press, 36-47, 1993.

[92] Harbaugh, R., (2003) "Skill Signaling, Prospect Theory, and Regret Theory" Claremont Colleges Working Paper No. 2002-03. Available at SSRN: http://ssrn.com/abstract=311409, Accessed on 18 May 2006

[93] Heuer, J. R., (1999) The psychology of intelligence analysis, Washington, DC: Center for the Study of Intelligence, Central Intelligence Agency, 1999.

[94] Hinds, P. J., and Pfeffer, J. (2003). Why Organizations Don't "Know What They Know": Cognitive and Motivational Factors Affecting the Transfer of Expertise. In Ackerman, M., Pipek, V. and Wulf, V. (eds.). Sharing Expertise: Beyond Knowledge Management, Cambridge: MIT Press

[95] Hinds, P., (1999) The curse of expertise: The effects of expertise and debiasing methods on predictions of novice performance. Journal of Experimental Psychology: Applied, 5, 205-221, 1999.

[96] Huber, G. (1991). Organizational Learning: The Contributing Processes and the Literatures. Organization Science 2:88-115.

[97] Huber, G. (1999). Experts in Organizations: The Power of Expertise. Academy of Business and Administrative Sciences (ABAS) International Conference, Barcelona, Spain,

[98] Iofciu, T., Diederich, J., Dolog, P., Balke, W., (2007) ExpertFOAF recommends experts, First International ExpertFinder Workshop, Berlin, Germany, January 16, 2007

[99] Jagdish, V. (2004) "A Signaling Theory of Managerial Turnover,", 2004. Available at SSRN: http://ssrn.com/abstract=613062

[100] Jensen, R. S., (1995) Pilot Judgment and Crew Resource Management. Cambridge: University Press.

[101] Jie, M., Karlapalem, K., and Lochovsky, F. (2000). A Multi-agent Framework for Expertise Location.

[102] Jin, R. and Dumais, S., (2001) Probabilistic Combination of Content and Links, SIGIR '01, September 9-12, 2001, New Orleans, LA, USA.

[103]   Jokinen, K. and Kanto, K., (2004) User Expertise Modelling and Adaptivity in a Speech-based E-mail System. In the Proceedings of the 42nd Annual Meeting of the Association of Computational Linguistics, ACL 2004, Barcelona, Spain.

[104]   Jones, Candace. (2003) Signaling Differentiation or Legitimacy?: How Architects Socially Construct Their Reputations for Expertise [Online] December 2003 http://www.ksg.harvard.edu/complexity/papers/Candance_Jones_and_Manev_abstract.pdf

[105]   Kahneman, D., Slovic, P., and Tversky, A. (Ed.). (1982). Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press.

[106]   Kaptelinin, V. (1996). Activity theory: Implications for human-computer interaction. In Nardi, B. A. (Ed.), Context and consciousness: Activity theory and human-computer interaction, 103-116. Cambridge, MA: The MIT Press.

[107]   Kautz H. and Selman B., (1998) "Creating Models of Real-World Communities with ReferralWeb", in Working Notes of the Workshop on Recommender Systems held in conjunction with AAAI-98, Madison, WI, 1998, pp. 58 – 59.

[108]   Kautz, B., Selman, B., and Shah, M. (1997). "The Hidden Web", AI Magazine, 18(2), 27-36.

[109]   Kleinberg, J., (1999) Authoritative Sources in a Hyperlinked Environment, Journal of the ACM, Vol. 46, No. 5, September 1999, pp. 604 –632.

[110]   Konchady, M. and D'Amore, R., (2002) Implementing a Smart Spider, Doctor Dobbs Journal, 2002, VOL. 27; Part 8, pages 62-68.

[111]   Kopec, J. A. (1995) Concepts of disability: the Activity Space Model, Social Science and Medicine, 40(5):649-56.

[112]   Korfhage, R.R. (1997). *Information Storage and Retrieval*. New York: John Wiley & Sons, Inc.

[113]   Krulwich B. and Burkey C., (1995) "ContactFinder: Extracting Indications of Expertise and Answering Questions with Referrals", in Working Notes of the 1995 Fall Symposium on Intelligent Knowledge Navigation and Retrieval, Cambridge, MA. Technical Report FS-95-03, The AAAI Press, 1995, pp. 85 – 91.

[114]   Krulwich, B., and Burkey, C., The ContactFinder agent: Answering bulletin board questions with referrals, in Proceeding of the Thirteenth National Conference on Artificial Intelligence (AAAI-96), AAAI Press, 1996.

[115] Kuutti, K. (1996) Activity theory as a potential framework for human computer interaction research. In Nardi, B. A. (Ed.), Context and consciousness: Activity theory and human-computer interaction, 17-44. Cambridge, MA: The MIT Press.

[116] Langer, E. J. and Imber, L. G. (1979). When practice makes imperfect: Debilitating effects of over learning. Journal of Personality and Social Psychology, 37, 2014-2024.

[117] Larres, M. P., Ballantine, J. A., and Whittington, M., (2003) Evaluating the validity of self-assessment: measuring computer literacy among entry-level undergraduates within accounting degrees at two UK universities. Accounting Education, 12 (2), 97-112.

[118] Lave, J. and Wenger, E. (1991), Situated Learning. Legitimate peripheral participation, Cambridge: University of Cambridge Press.

[119] Leibowitz, J. (2001). Knowledge Management: Learning from Knowledge Engineering. Boca Raton, FL: CRC Press.

[120] Leonard, D. and Swap, W., (2005), Deep Smarts: How to Cultivate and Transfer Enduring Business Wisdom. Boston, MA: Harvard Business School Press (2005), pp. 203–226.

[121] Leont'ev, A. N. (1974): 'The problem of activity in psychology', Soviet Psychology, vol. 13, pp. 4-33.

[122] Lesser, E., and Storck, J., (2001) Communities of practice and organizational performance. IBM Systems Journal, 40(4), 831-841, 2001.

[123] Li, J., Boley, H., Virendrakumar, C., Mei, J. (2006) Expert Finding for eCollaboration Using FOAF with RuleML Rules, Montreal Conference on eTechnologies (MCeTech).

[124] Lifantsev, M. (2000) Voting model for ranking Web pages. In Peter Graham and Muthucumaru Maheswaran, editors, Proceedings of the International Conference on Internet Computing, pages 143–148, Las Vegas, Nevada, June 2000.

[125] Lindstaedt S, (1996), Towards Organizational Learning; Growing Group Memories in the Workplace. In Tauber M (Ed) CHI 96 pp 53-54.

[126] Liu, X., Croft, W. Bruce, and Koll, M., (2005) Finding Experts in Community-based Question-Answering Services, Proceedings of ACM CIKM,. Bremen, Germany.

[127] Lukose, R., Adar, E., Tyler, J., and Sengupta, C., (2003) SHOCK: Communicating with Computational Messages and Automatic Private Profiles, The Twelfth International World Wide Web Conference (2003)

[128]   MacDonald, C., He, B., Plachouras, V., Ounis, I.. (2005) University of Glasgow at TREC 2005: Experiments in Terabyte and Enterprise Tracks with Terrier. Proceedings of the Fourteenth Text REtrieval Conference (TREC 2005), Voorhees, E. M., and Buckland, L. P. , editors. Gaithersburg, MD.

[129]   Maglio, P. and Matlock, T. (1999). The conceptual structure of information space. In A. Munro, K. Hook, & D. Benyon (Eds.), Social navigation of information space, London: Springer-Verlag, pp 155-173.

[130]   Manmatha, R., Rath, T., and Feng, F., (2001) Modeling Score Distributions for Combining the Outputs of Search Engines, Proceedings of SIGIR '01, September 9-12, New Orleans LA.

[131]   Maron, M.E., Curry, S., and Thompson, P., (1986) "An Inductive Search System: Theory, Design and Implementation", IEEE Transaction on Systems, Man and Cybernetics, vol. SMC-16, No. 1, pp 21-28, January/February 1986.

[132]   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.

[133]   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.

[134]   Maybury, M., and D'Amore, R. (2001) Discovery of Expert Communities and Social Interests. AAAI Workshop on Socially Intelligent Agents- The Human in the Loop.

[135]   Maybury, M., D'Amore, R, and House, D. December (2001) Expert Finding for Collaborative Virtual Environments. Communications of the ACM 44(12): 55- 56. In Ragusa, J. and Bochenek, G. (eds). Special Section on Collaboration Virtual Design Environments, ACM .

[136]   Maybury, M., D'Amore, R., and House, D. (2000) Automating expert finding. Intern. J. Technology Research Manage. 43, 6. (Nov-Dec 2000), 12-15.

[137]   Maybury, M., D'Amore, R., and House, D. (2003). Automated Discovery and Mapping of Expertise. In Ackerman, M., Pipek, V. and Wulf, V. (eds.). Sharing Expertise: Beyond Knowledge Management, Cambridge: MIT Press

[138]   Maynard Smith, J., and Harper, D., (2003) Animal Signals. Oxford University Press, Oxford.

[139]   McDonald D.W., and Ackerman M.S., (1998) Just Talk to Me: A Field Study of Expertise Location, Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work (CSCW'98), Seattle, WA: ACM Press, 1998, pp. 315-324.

[140]   McGovern, J., Sims, O., Jain, A. and Little, M. (2006), Enterprise Service Oriented Architectures: Concepts, Challenges, Recommendations, Springer.

[141]   McLean, A., Samaratunge S., Vercoustre A-M., and Wu, M., (2003) Enterprise PeopleFinder: Combining evidences from Web pages and Corporate data", The 8th Australasian Document Computing Conference (ADCS'03), 2003.

[142]   Montague, M. and Aslam, J. A., (2002) Condorcet fusion for improved retrieval. In Proc. 11th CIKM, pages 538--548, 2002.

[143]   Morville, P., (2005) Ambient findability: what we find changes who we become. Cambridge: O'Reilly.

[144]   Mynatt, E. D., Adler, A., Ito, M., Linde, C. and O'Day, V. L., (1999) The net communities of SeniorNet. Proceedings of ECSCW 1999.

[145]   Nardi, B., Editor. (1996). Context and Consciousness: Activity Theory and Human-Computer Interaction. Cambridge: MIT Press.

[146]   Nardi, B., Whittaker, S., Isaacs, E., Creech, M., Johnson, J., and Hainsworth, J. (2002). ContactMap: Integrating communication and information through visualizing personal social networks. Communications of the Association for Computing Machinery. April. Pp. 89-95.

[147]   Nardi, Whittaker, and Schwarz (2002) Nardi, B., Whittaker, S. and Schwarz, H. (2002). NetWORKers and their activity in intensional networks. The Journal of Computer-supported Cooperative Work 11: 205-242.

[148]   Neiman, F. D. (1998). Conspicuous consumption as wasteful advertising: a Darwinian perspective on spatial patterns in the Classic Maya terminal monument dates. In: Barton CM, Clark GA, editors. Rediscovering Darwin: evolutionary theory and archaeological explanation. Washington D. C.: Archaeological Papers of the American Anthropological Association, No. 7. p 267–290.

[149]   Ogata, H., Fukui, T., Yano, Y. (1999) SocialPathFinder: Computer Supported Exploration of Social Networks on WWW. ICCE 99.

[150]   Orasanu, J., and Connolly, T. (1993). The reinvention of decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), Decision making in action: Models and methods, (pp. 3-20). Norwood, NJ: Ablex Publishing

[151]   Orr, Julian E. (1996) Talking about machines: An ethnography of a modern job. Ithaca, NY: Cornell University Press

[152]   Peters, C., and Borri, F., (Eds.) (2004) Results of the CLEF 2004 Cross-Language System Evaluation Campaign, Working Notes for the CLEF 2004 Workshop 15-17 September, Bath, UK

[153]   Pipek, V., Hinrichs, J., and Wulf, V. (2003), Sharing Expertise: Challenges for Technical Support, in M. Ackerman, V. Pipek, and V. Wulf (eds.): Sharing Expertise - Beyond Knowledge Management, MIT Press, Cambridge, 2003, pp. 111-136.

[154]   Reichling, T., and Veith, M. (2005) Expertise sharing in a heterogeneous organizational environment. In Proceedings of the 9th European Conference on Computer-Supported Cooperative Work, 18 – 22 September 2005, Paris (Dordrecht: Springer), pp. 325 – 345.

[155]   Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E., (2000). Reputation systems. Communications of the ACM, 43(12):45–48.

[156]   Rheingold, H (1993) The Virtual Community. Reading: Addison-Wesley Publishing Company

[157]   Rickleffs, R. E., Ecology (3rd Edition), W. H. Freeman & Co Ltd. New York. 1990.

[158]   Rosenbloom, P. S. and Newell, A. (1986). The chunking of goal hierarchies: A generalized model of practice. In R. S. Michalski, J. G. Carbonell and T. M. Mitchell (Eds.), Machine Learning: An Artificial Intelligence Approach, Volume II (pp. 247-288). Los Altos, CA: Morgan Kaufmann Publishers, Inc.

[159]   Ru, Z., Chen, Y., Xu, W., Guo, J., TREC 2005 Enterprise Track Experiments at BUPT, In E. M. Voorhees and L. P. Buckland, editors. Proceedings of the Fourteenth Text REtrieval Conference (TREC 2005), Gaithersburg, MD, November 2004.

[160]   Saari D. G. (1999). 'Explaining all three-alternative voting outcomes', *Journal of Economic Theory* 87 (2): 313-355.

[161]   Salton, G. and McGill, M. J. (1983) Introduction to Modern Information Retrieval. New York: McGraw-Hill.

[162]   Salton, G., (1971) Cluster search strategies and the optimization of retrieval effectiveness. In G. Salton, editor, The SMART Retrieval System, pages 223-242. Prentice- Hall, Englewood Cliffs N J, 1971.

[163]   Sanderson, M. and Joho, H., (2004) Forming test collections with no system pooling. In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 33–40, 2004.

[164]   Sastry, N., (2006) Expertise in Social Networks Technical Report #: 06-11, In Proceedings of the First IEEE Workshop on Hot Topics in Web Systems and Technologies (HotWeb 2006), Boston, MA, USA, November 2006. IEEE Computer Society.

[165]   Schlichter, J., Koch, M. and Chengmao, X. (1998). Awareness - The Common Lin Between Groupware and Community Support Systems. In Ishida, T. (Ed.), Community Computing and Support Systems - Social Interactions in Networked Communities. Berlin: Springer-Verlag.

[166]   Schwartz, M. F. and Wood, D. C. M., (1993) Discovering shared interests using graph analysis. Communications of the ACM, 36(8):78-80, 1993.

[167]   Selznik, P. (1996). In search of community. Rooted in the Land: Essays on Community and Place. W. Vitek and W. Jackson. New Haven, Yale University Press: 195-203

[168]   Shah, C and Croft, W. Bruce, (2004) Evaluating High Accuracy Retrieval Techniques. SIGIR 2004. Sheffield, UK. July 2004.

[169]   Shanteau, J. (1992). Competence in experts: The role of task characteristics. Organizational Behavior and Human Decision Processes, 53, 252-266.

[170]   Shanteau, J., (1993) Discussion of expertise in auditing. Auditing: A Journal of Practice and Theory, 12, 51-56.

[171]   Shanteau, J., Weiss, D. J., Thomas, R. P., and Pounds, J. (2003). How can you tell if someone is an expert? Performance-based assessment of expertise. In Schneider, S. L., and Shanteau, J. (Eds.), Emerging perspectives on judgment and decision research. (pp. 620-639). Cambridge, UK: Cambridge University Press.

[172]   Shanteau, J., Weiss, D. J., Thomas, R., and Pounds, J. (2002). Performance-based assessment of expertise: How can you tell if someone is an expert? European Journal of Operational Research, 136, 253-263.

[173]   Sheehan, K. and Hoy, M. G. (1997), E-mail Surveys: Response Patterns, Process and Potential. Presented at the 1997 American Academy of Advertising Conference, April 7, 1997, in St. Louis, MO.

[174]   Shrauger, J. S. and T. M. Osberg. (1981). The relative accuracy of self-predictions and judgments by others in psychological assessment. Psychological Bulletin 90(2):322-351.

[175]   Simon, H. A., and Chase, W. G. (1973). Skill in chess. American Scientist, 61, 394-403

[176]   Simon, H. A., and Gilmartin, K. (1973). A simulation of memory for chess positions. Cognitive Psychology, 5, 29-46

[177]   Small, C., and Zoracki, M., (2000) "Potential Role of Knowledge Management and Agent Technology", Internal MITRE Briefing. July 2000

[178]   Snijders, T. (1992) 'Estimation on the basis of snowball samples: how to weight', Bulletin Methodologie Sociologique, vol. 36, 59-70.

[179]   Soboroff, I. and Robertson, S. (July 2003), Building a Filtering Test Collection for TREC 2002, Proceedings of the 26th Annual International Conference on Research and Development in Information Retrieval (SIGIR 2003), Toronto, Ontario, Canada.

[180]   Soboroff, I., Nicholas, C., and Cahan, P., (2001) Ranking Retrieval Systems without Relevance Judgments Proceedings of the 24th Annual International Conference on Research and Development in Information Retrieval (SIGIR 2001), New Orleans, LA.

[181]   Soboroff, I., and Harmon, D., Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), pages 105–112, Vancouver, October 2005. 2005 Association for Computational Linguistics

[182]   Sosis, R., and E. Bressler (2003), Cooperation and Commune Longevity: A Test of the Costly Signaling Theory of Religion,. Cross-Cultural Research, 37:211–239.

[183]   Sparck Jones, K. and Van Rijsbergen, C. J., (1976). Information retrieval test collections. J. Doc. 32, 1 (Mar.), 59-75.

[184]   Stewart, T. R., Roebber, P. J., and Bosart, L. F., (1997): The importance of the task in analyzing expert judgment. Organizational Behavior and Human Decision Processes, 69.

[185]   Streeter, L. and 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.

[186]   Surowiecki, J. (2004). The Wisdom of Crowds: Why the Many Are Smarter than the Few, and How Collective Wisdom Shapes Business, Economies, Societies, and Nations. New York: Doubleday. 296 pages. ISBN 0-385-50386-5.

[187]   Terveen, L. G. and D. W. McDonald (2005) Social Matching: A Framework and Research Agenda. ACM Transactions on Computer-Human Interaction (ToCHI). 12 (3), pp. 401-434.

[188]   Tukey, W., (1977), Exploratory data analysis, Addison-Wesley, Reading, Mass.

[189]   Tversky, A. and Kahneman, D. (1973) Availability: A heuristic for judging frequency and probability. Cognitive Psychology 5: 207-232.

[190]   Vandeventer, S. (1997). Expert Behaviour among outstanding Videogame-playing Children. Unpublished Dissertation submitted in partial fulfillment of the requirements of the Doctor of Philosophy (Curriculum and Instruction). University of South Florida., Florida.

[191]   Veblen, Thorstein. (1899). The theory of the leisure class. New York: Macmillan.

[192]   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.

[193]   Voorhees, E. M. (2003). Overview of the TREC 2003 Question Answering Track. In Proceedings of the Twelfth Text REtrieval Conference (TREC 2003).

[194]   Voorhees, E. M., and Tice, D. M. (2000). Building a Question Answering Test Collection. In Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2000).

[195]   Voorhees, E. M., Gupta, N. K., Johnson-Laird, B., (1995) Learning collection fusion strategies, Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval, p.172-179, July 09-13, 1995, Seattle, Washington, United States

[196]   Wang, J., Chen, Z., Tao, L., Ma, W., Wenyin, L. (2002), Ranking User's Relevance to a Topic through Link Analysis on Web Logs, WIDM'02, November 8, 2002, McLean, Virginia, USA.

[197]   Wasserman, S. and Faust, K., (1994) *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

[198]   White, S., and Smyth, P., (2003) Algorithms for Discovering Relative Importance In Graphs. *Proceedings of Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington D.C., 2003.

[199]   Whittaker, S., Issacs, E., and O'Day, V. (1997). Widening the net. Workshop report on the theory and practice of physical and network communities. SIGCHI Bulletin, 29(3), 27-30.

[200]   Whittaker, S., Jones, Q., Nardi, B., Creech, M., Terveen, L., Isaacs, E., and Hainsworth, J. (2004). ContactMap: Organizing Communication in a Social Desktop. ACM Transactions on Computer Human Interaction, 11, 445 – 471

[201]   Williamson, S. and Wright, R. (1994), "Barter and monetary exchange under private information," American Economic Review, 84, 104-123.

[202]   Won, M. and Pipek, V., (2003) Peripheral Expertise Awareness – Sharing Knowledge on Knowledge. In: Proceedings i-Know, Graz, Austria, 2003, pp. 155-160.

[203]   Wooldridge, M. and Jennings, N. R., (1994) Formalizing the Cooperative Problem Solving Process , 13th International Workshop on Distributed Artificial Intelligence (IWDAI-94), Lake Quinhalt, WA, July 1994, 403-417.

[204]   Yao, C., Peng, B., He, J., and Yang, Z., (2005) CNDS Expert Finding System for TREC 2005 In Proceedings of the Fourteenth Text REtrieval Conference (TREC 2005), Voorhees, E. M., and Buckland, L. P. , editors. Gaithersburg, MD.

[205]   Yimam, D. (1999). Expert Finding Systems for Organizations: Domain Analysis and the DEMOIR approach. ECSCW 99 Workshop: Beyond Knowledge Management: Managing Expertise.

[206]   Yimam-Seid, D. and A. Kobsa (2003): Expert Finding Systems for Organizations: Problem and Domain Analysis and the DEMOIR Approach. Journal of Organizational Computing and Electronic Commerce 13(1), 1-24.

[207]   Yu, B. and Singh, M. P. (2002).   Searching Social Networks. AAMAS 2002 Workshop on Regulated Agent-Based Social Systems: Theories and Applications (RASTA).

[208]   Zahavi, A., (1975) Mate selection - selection for a handicap. Journal of Theoretical Biology 53, no. 1: 205-214.

[209]   Zahavi, A., and Zahavi, A. (1997). The Handicap Principle: A Missing Piece of Darwin's Puzzle. New York: Oxford University Press.

[210]   Zhu, W., Song, M., and Allen, R. B., (2005) TREC 2005 Enterprise Track Results from Drexel. In Proceedings of the 14th Text REtrieval Conference (TREC 2005), Gaithersburg, MD, USA, 2005.

12 Appendix A: *Expert Locator* Survey Form

Tool Evaluation Survey

This short survey was designed to support an early evaluation of a social information retrieval prototype. Your feedback will be used to assess our progress and validate system performance. Your responses will be kept private. All data included in any reports will be anonymized.

You have been identified as having some knowledge of the domain of **bioinformatics**. Please take a few minutes of your time to answer the following questions regarding your familiarity with this domain. Keep in mind that we are evaluating the system - we are not evaluating you.

How would you <u>define your relationship</u> to this domain? (There may be more than one applicable answer.)

☐ **practitioner** (involved in relevant research or sponsor applications)

☐ **broker** (know something about the domain, able to point people to sources of expertise)

☐ **not involved**

If you have identified yourself as **not involved**, please skip the rest of these questions and submit this form.

Not counting college education, how many <u>years of work experience</u> do you have in this domain?

< 1	1-3	3-5	5+	N/A
☉	☐	☐	☐	☐

How many <u>publications</u> have you (co)authored in this domain?

0	1	2-5	6-10	10+
◉	☐	☐	☐	☐

Please <u>rate your knowledge</u> of this domain from 1 (don't know much or just learning) to 5 (highly knowledgeable).

1	2	3	4	5	not sure
☐	☐	☐	☐	☐	◉

don't know much   **highly knowledgeable**

The system has also identified some of these MITRE employees as having knowledge of the domain. Help us <u>rate the system's performance</u> by agreeing or disagreeing with each assessment.

	strongly agree	agree	not sure	disagree	strongly disagree	don't know this person
Smith, John S.	☐	☐	☐	☐	☐	◉
Jones, Mary K.	☐	☐	☐	☐	☐	◉
Dobby, J. K.	☐	☐	☐	☐	☐	◉

Please list the names of any other MITRE employees you consider to be knowledgeable in this domain. (Please separate names with a semicolon.)

Additional comments, if any:

Thanks for your feedback!

Submit Form Bottom of Form

12.1.1.1.1 *Help* | *Questions?*

13   Appendix B: Selected Precision Results

Precision scores for current queries are presented here in slightly more detail. In each table *R-known* are the relevant known and *NR-known* are non-relevant known. *RELret* are the number of known relevant actually retrieved by *Expert Locator* within the *R-known* ranks.   The last column provides the second performance measure, *Pr (Top 5),* which is the probability that a known relevant is in the top 5 ranks. As with R-precision, *Pr (Top 5)* is computed over three populations, authorities, hubs, and the third case which is the inclusive OR of authorities and hubs.   For completeness, the number of known non-relevant retrieved, *NR-ret*, is computed; this is available to support a finer-grained assessment of system errors. *R-precision* and *Pr (Top 5)* are computed across all 36 queries with the mean *R-precision* and mean *Pr (Top 5)* used as system performance measures.

Authorities								
Query	r-known	nr-known	rel-ret	r-precision	Pr(Top 5)	NR-Ret	UNK's	%UNK
Air Traffic Control	23	6	4	0.174	0.6	2	17	0.739
Bayesian Networks	13	22	6	0.462	0.8	1	6	0.462
Biocomputing	16	19	3	0.188	0.6	7	6	0.375
Biometrics	21	6	8	0.381	0.6	4	9	0.429
Brain Mapping	12	9	7	0.583	0.6	2	3	0.250
Chemical Warfare	9	25	3	0.333	0.4	5	1	0.111
Complex Adaptive Systems	47	18	13	0.277	1.0	6	28	0.596
Geospatial Mapping	34	13	4	0.118	0.8	4	26	0.765
Global Position System	14	13	1	0.071	0.2	8	5	0.357
Grid Computing	18	5	9	0.500	0.8	2	7	0.389
Homeland Security	47	14	12	0.255	0.4	4	31	0.660
Human Computer Interaction	56	9	20	0.357	0.8	7	29	0.518
Information Retrieval	58	7	27	0.466	1.0	7	24	0.414
Insider Threat	20	6	10	0.500	0.8	3	7	0.350
J2EE	30	12	9	0.300	0.6	4	17	0.567
Logistics	20	12	3	0.150	0.4	3	14	0.700
Nanotechnology	8	7	6	0.750	0.8	1	1	0.125
Network Protocols	38	25	8	0.211	0.6	10	20	0.526
Operationsl Research	25	9	6	0.240	0.4	5	14	0.560
Robotics	8	21	6	0.750	0.6	2	0	0.000
Social Network Analysis	12	14	4	0.333	0.6	5	3	0.250
TOTAL	529	272	169	0.319	0.639	92	268	0.507

Table 13-1: Authority Scores

Hubs								
Query	r-known	nr-known	rel-ret	r-precision	Pr(Top 5)	NR-Ret	UNK's	%UNK
Air Traffic Control	10	22	4	0.400	0.4	1	5	0.500
Bayesian Networks	11	24	6	0.545	0.6	0	5	0.455
Biocomputing	9	26	3	0.333	0.6	4	2	0.222
Biometrics	5	22	2	0.400	0.4	3	0	0.000
Brain Mapping	8	13	3	0.375	0.6	5	0	0.000
Chemical Warfare	4	30	2	0.500	0.4	2	0	0.000
Complex Adaptive Systems	22	43	12	0.545	1.0	2	8	0.364
Geospatial Mapping	6	41	0	0.000	0.0	5	1	0.167
Global Position System	7	20	4	0.571	0.8	3	0	0.000
Grid Computing	9	14	3	0.333	0.6	3	3	0.333
Homeland Security	16	45	5	0.313	0.4	5	6	0.375
Human Computer Interaction	27	38	12	0.444	0.6	5	10	0.370
Information Retrieval	24	41	19	0.792	1	5	0	0.000
Insider Threat	7	19	2	0.286	0.4	4	1	0.143
J2EE	10	32	3	0.300	0.4	4	3	0.300
Logistics	6	32	2	0.333	0.4	2	2	0.333
Nanotechnology	5	10	1	0.200	0.2	4	0	0.000
Network Protocols	10	53	4	0.400	0.6	6	0	0.000
Operationsl Research	9	255	2	0.222	0.4	3	4	0.444
Robotics	8	22	5	0.625	0.8	3	0	0.000
Social Network Analysis	11	15	7	0.636	0.8	2	2	0.182
TOTAL	224	817	101	0.451	0.543	71	52	0.232

Table 13-2: Hub Scores

Hubs or Authorities							
Query	r-known	nr-known	rel-ret	r-precision	Pr(Top 5)	NR-Ret	UNK's	%UNK
Air Traffic Control	29	9	6	0.207	0.6	0	23	0.793
Bayesian Networks	19	16	12	0.632	0.8	1	6	0.316
Biocomputing	20	15	6	0.300	0.8	5	9	0.450
Biometrics	20	15	6	0.300	0.8	1	13	0.650
Brain Mapping	17	4	9	0.529	0.8	1	7	0.412
Chemical Warfare	10	24	4	0.400	0.6	2	4	0.400
Complex Adaptive Systems	54	11	18	0.333	1.0	1	35	0.648
Geospatial Mapping	38	9	6	0.158	0.8	1	31	0.816
Global Position System	20	9	6	0.300	0.8	3	11	0.550
Grid Computing	20	3	11	0.550	0.8	0	9	0.450
Homeland Security	51	10	14	0.275	0.6	2	35	0.686
Human Computer Interaction	60	5	24	0.400	0.8	4	32	0.533
Information Retrieval	59	6	27	0.458	1	4	28	0.475
Insider Threat	23	3	13	0.565	0.8	0	10	0.435
J2EE	34	8	12	0.353	0.8	1	21	0.618
Logistics	27	8	7	0.259	0.8	1	19	0.704
Nanotechnology	11	4	8	0.727	0.8	1	2	0.182
Network Protocols	44	19	12	0.273	1	3	29	0.659
Operationsl Research	30	4	11	0.367	0.6	1	18	0.600
Robotics	13	17	8	0.615	0.8	1	4	0.308
Social Network Analysis	17	9	8	0.471	0.8	2	7	0.412
TOTAL	**616**	**208**	**228**	**0.370**	**0.790**	**35**	**353**	**0.573**

Table 13-3:   Combined Hubs and Authority Score

14 Appendix C: Selective Activity Space Descriptions

The activity spaces described in Chapter 5 are more fully described here. This includes AS's selected for use in the *Expertise Locator* model as well as some that have not been integrated but by through their expanded descriptions provide more context for the overall AS selection. The potential evolution of these spaces relative to changes in the host environment is discussed where appropriate.

14.1 Activity Space Taxonomy

The activity space descriptions and selection process carried out in Chapter 5, are summarized in Table 14-1; with the final AS selections shown in Table 14-2. AS's <u>not</u> selected for broader discussion are **bolded**. \

Activity Space Classification	MII: Activity Spaces	Expertise Aspect
Organization/Personal	Public Share, Private Share, Blogs, About-Me, **E-mail, Instant Messaging (IM)**	Personal spaces used to convey user interests, knowledge, or expertise. Each personal space is linked to user's home organization (e.g., department).
Corporate Technical Teams	Technology Area Teams (TATs), Skills Clusters, The Hotline, MITRE Repository of Knowledge (MRoK)	Team-based spaces formed around corporate teams and related to specific expertise areas or expertise services
Projects	Project Page, Project Share, SourceForge	Team-based workspaces set up to organize, store, and share project work consistent with access constraints (e.g., privacy or security)
Community	Sharepoint, ListServs, Technical Exchange Meetings (TEMs)	Collaborative spaces that support multi-user communication and information sharing.

Table 14-1: Activity Space Classes and Instances within the MII

Selected Prototype Activity Spaces	Corporate Space
Project	**Project Page, Project Share, TATs**
ListServs	**ListServs**
Organization/Person	**Public Share**, About-Me

Table 14-2: Prototype Activity Spaces Built Mapped to Corporate Spaces

14.2 Expanded Activity Space Descriptions

14.2.1 Public Share Folders

MITRE staff members are associated with a "home" organization (department). Each department member is associated with a personal space used to publish and share information. This department-centric personal space is viewed as distinct from a staff member's projects, corporate technical teams, and their involvement in various communities.

To promote knowledge sharing, each employee and contractor (with system access) has a Public Share folder. Public Share folders are typically used as a type of online storage or personal information space. Users can drag-and-drop documents into their folders for sharing and at the same time publish documents to the corporate collection. Publishing is seamlessly handled by the search engine; Share Folders are "visited" by the search engine spider, objects are cached and indexed for retrieval. Users have options to add metadata that may facilitate retrieval and improve performance for applications that require author-identification. For example, the system provides a capability for a user to enter author name and optionally other topic descriptors prior to publishing.

A Public Share folder can be hierarchically organized into subfolders as shown in Figure 14-1. As shown below, some documents are stored at the root with others assigned into separate subfolders. Figure 14-2 shows the Public Share artifact distribution for the entire enterprise (November 2004.) The average is approximately 26 documents per person. Roughly 70% of the Public Share folders have at least 5 documents and more than 50% have more than 10 items. Closer inspection of actual Public Share folders, suggests that highly skilled knowledge workers, for example, Technology Area Team (TAT) members, have considerably more items per Folder than the average worker. This provides some evidence at least that most "public" experts will exploit Share Folders for information sharing and advertising expertise. There are instances, however, where work sensitivity precludes open sharing and this may affect Public Share coverage on key expertise areas.

Figure 14-1:   Public Share Folder
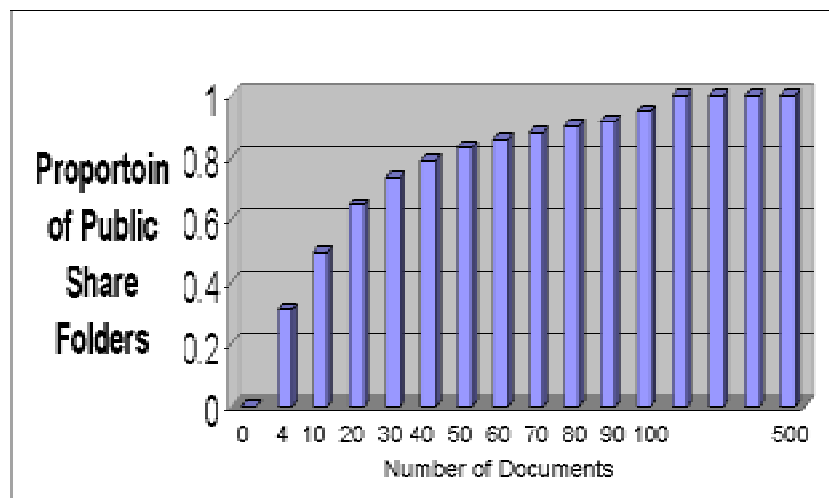
Figure 14-2: Cumulative Distribution of Enterprise Public Share Folders (~4500 actors)

Overall, Public Share folders provide a rich repository from which to glean expertise. The usefulness of Public Share folder data for expertise detection was demonstrated earlier using the XperNet system, developed by D'Amore in Maybury, D'Amore, and House, (2002). XperNet used document clustering and person attribute data to extract expertise areas.

14.2.2   About-Me Folder

The About-Me folder can be used to publish professional information about a user's skills and experiences. The About-Me folders can be written to only by their owners, and require MITRE domain authentication. Other users can view files in an About-Me folder as well as copy files from a person's About Me folder. The About-Me folder has significant potential for use in an automated expert finder system in part because it is not intended overtly for expert finding and since they may have other uses, they may more likely be kept current. About-Me is becoming increasingly important as an informal version of the traditional resume and has been used to document yearly performance; necessary to support annual performance reviews. Resumes are more problematic in terms of their update and currency, while About-Me folders may be easier to maintain and more relevant to documenting finer grained work experience. In addition, About-Me addresses the information needs of an internal (corporate) audience in contrast with resumes which are often for use with external organizations.

About-Me is semi-structured; users may use "fields" or "tags" to denote certain entries in their description (for example: Programming Languages) but this is not a requirement. Users can simply enter text describing work performed or special expertise. Here a user may provide information related to their skills and project experience indirectly through items submitted to the folder.   Instantiating an About-Me is in effect a form of "registration" process in which the user can effectively *signal* their skill areas to a wider audience using a corporately supported mechanism.

14.2.3   Blogs

"A Blog is a web page made up of usually short, frequently updated posts that are arranged chronologically—like in a "what's-new*"* page or a journal. The content and purposes of Blogs varies greatly—from links and commentary about other web sites, to news about a company/person/idea, to diaries, photos, poetry, mini-essays, project updates, even fiction.[132]" Blogs @ MITRE is an interactive content management system that provides a simple way for all

---

[132] www.blog.com. Accessed on November 18 2005

MITRE employees and contractors to post information regarding their individual or project work.   In a narrow sense a Blog is a document list in reverse chronological order that, for the purposes of this research, can be used to infer an author's interests or, potentially, expertise. However, somewhat like ListServs, Blogs encourage interaction with others; for example, Blog visitors can respond to particular posts creating an exchange forum centered on the Blog owner's interest and with TrackBack[133] a Blogger can notify another when posting something of mutual interest. Blogs (or more appropriately Blog software) provide content management support (e.g., for archiving posts) and search. Publication support is getting increasingly sophisticated to include post scheduling for publishing at pre-determined times, and image handling without the need for special software. Blogs (including Blogs@MITRE) are providing increasingly more support for groups and even communities that go beyond instant publication, easy file sharing using attachments, and subscription services supporting email notification. While the immediate focus here is on Blogs as personal spaces, the natural evolution is towards increased support for groups and communities.

Blogs are a potentially rich space for ferreting out expertise. While not as widely established at MITRE as ListServs they are becoming increasingly more prevalent as a means for individuals and groups to share information. While there are technical differences between ListServs and Blogs the main interest here is in publication control, content management, and interaction.

- **Publication Control**: The Blogger (individual or small group) essentially dictates Blog topics and publication schedule; while ListServs are under multi-author control providing group, decentralized publication. Blogs are inherently more suited for reflecting individual views and ListServs are more aligned with group interests. However, individuals or groups can be served by either.

- **Content Management**: Blogs are generally more structured than a typical ListServ in that posts may be grouped consistent with defined topics. This is in contrast with ListServs that typically handle domain or thematic variation by defining relatively homogeneous forums. ListServs are like *attractors* in which users with common interests group. As such, ListServs may be closer to a faceted

---

[133] www.movabletype.org. Accessed on November 20 2005

(single level categorization) while Blogs may be more hierarchically organized on a variety of topics. Underlying this framework, Blogs like ListServs provide tools for archiving and searching posts.

- **Interaction**: While Blogs were originally designed as personal spaces, they are increasingly becoming more group or community oriented. Comments and various connection protocols like TrackBack increase support for community building. There are significant implications here for expertise detection. In particular, an expertise detection scheme may exploit individual Blogs to infer expertise but may exploit links between Blogs in detecting expert communities. The main interest here is in using Blogs to ferret out expertise networks; communities of practice involving a set of experts that may be coupled by common knowledge base and experience as well as by overlapping work relationships.

14.2.4   Skill Cluster Groups

The objective of Skill Cluster Groups is to keep MITRE personnel abreast of technology developments. Cluster Groups are organized around various special interests and skill areas and are committed to disseminating technical information and providing referrals to outside and internal experts. The main impetus behind Clusters Groups is to support MITRE's core business and, as a result, Cluster Groups as relatively static domains. Of course, over time, shifts in work focus, emergence of new technologies and evolving sponsor needs results in gaps. Many uncovered areas are being addressed increasingly by various communities-of-interest such as those supported by ListServs and Community Share. Therefore, Cluster Groups complement other forums that may more rapidly adapt to changes in technology interests or the emergence of special problems. More generally, this suggests the use of multiple (complementary) AS's in an expert finder application; one or a few activity spaces may not cover well both the stable and emerging areas of expertise critical to the enterprise.

Cluster Groups, Table 14-3, are fairly diverse, spanning a number of disciplines from programming languages, to analysis techniques, to business coverage on whole industries. The MATLAB Cluster Group, for example, is typical of the specialty areas found.

Ada Programming Language	Image Processing and Visualization	MITRE Washington Macintosh Users Group
Artificial Intelligence and Decision Support Systems	Information Systems Architectures (ISA)	Natural Computation
Civil Aviation Operations Cluster Group	Instructional Technology Working Group	Operations Research and Mathematical Sciences (ORMS)
Data Mining Group	Java Cluster Group	Perl Prototyping and Programming
Database Management Systems (DBMS)	Knowledge Management	Push Technology
Digital Signal Processing	Language Technology Cluster Group	Reuse and Domain Engineering
Geographic Information Systems (GIS)	**MATLAB**	Risk and Reliability Analysis

Table 14-3: MITRE Skill Cluster Groups

As shown in Figure 14-3, the MATLAB Cluster Group is focused on the MATLAB tool[134] and how it is used in various fields. The purpose of the group is to facilitate collaboration amongst members so as to increase their proficiency in MATLAB use and better understand how MATLAB can be used to solve particular problems. Beyond this the group provides expertise to individual staff, project teams, and sponsors. From an enterprise modeling perspective, cluster groups like MATLAB, provide an activity space or work context that may be "mined" by expert-finder evidence collection and indexing tools to capture expertise within each domain. Spaces like MATLAB are culled here as activity spaces in that the work domain is formally bounded (corporate mandate), has membership, and associated activities. Therefore it is an activity space providing both artifact and social evidence of expertise.

---

[134] http://www.mathworks.com/ Accessed on October 22 2005
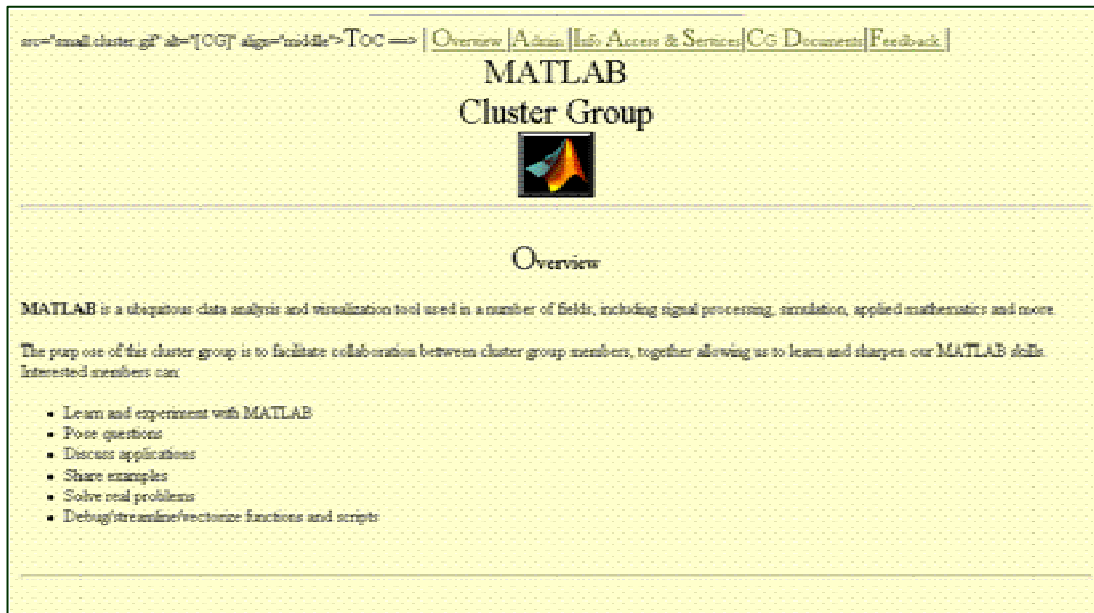
Figure 14-3: MATLAB Cluster Group Overview

Cluster group pages have structure that can be used to identify key group roles and membership-supporting services as shown in Figure 14-4. For example, members[135] have a simple script they can run that allows them to archive a document (with author attribution) to the MATLAB space. Group e-mail can be used to distribute information of general interest. In addition, MITRE staff (not just MATLAB members) can pose questions to the group.

Figure 14-4: MATLAB Administrative and Organizational Services

Users with particular expertise and who have group roles in terms of content management and interaction with other MITRE staff are listed as shown in Figure 14-5. This association of person to expertise sub-area provides additional sources of expertise evidence that may be used by an expert locator tool. In addition, other internal and external resources are listed below and these links can be used to identify the scope of the MATLAB group.

Figure 14-5: MATLAB Information and Services Available

Users are also provided with various resources to include tutorials and access to special applications, Figure 14-6.

---

[135] Project members' names are masked out consistent with corporate policy.

Figure 14-6: Group Resources

The relative homogeneity of Cluster Groups as to topic areas covered allows for hybrid expertise characterization schemes where (similar to cluster retrieval techniques used in standard information retrieval, Salton (1971), Salton and McGill (1983), whole memberships may be retrieved based on matching evidence at the Cluster Group level. Group signature (cluster) matching could complement search strategies that exploit authorship information and stored questions-answer pairs to locate individual experts. Cluster Groups are not currently indexed for retrieval and as discussed in Chapter 5 are not included in the current *Expert Locator* implementation.

14.2.5 *Technology Area Teams*

Technology Area Teams (TATs) are part of MITRE's Technology Program (MTP). The MTP is led by MITRE's Chief Technology Officer (CTO); and supported by Chief Engineers (CE's) from each operational center. TATs differ from Cluster Groups in that they have a formal organizational role to provide the CTO and the CE's with assessments of internal and external research on an ongoing basis and across thirteen major technology areas. Each TAT consists of technical experts from across MITRE's operational centers. TATs prepare forward looking assessments on current and emerging technologies, support proposal review during the MTP

research funding competition, and generally provide support to staff members and sponsors in areas related to their expertise. TATs members are nominated by senior management and technologists based on their expertise and work accomplishments. Most have prior or ongoing research experience. As such, TATs represent expertise areas that staff members can access but they also may be used by expert finding tools to identify expertise areas and specific experts. Current TAT areas are included in Table 14-4.

Technology Area Team	Description
Biotechnology	The Biotechnology TAT focuses on biomedical research as it intersects with information technology, security, national intelligence, and defense. This includes biomedical and neuroscience informatics, computational biology and biologically inspired computation, biosecurity and biodefense, and biosensing (including both sensing of biological agents and biologically-based sensors).
Communications and Networks	Communications covers LAN and WAN network protocols, system planning, management, traffic analysis, wireless technologies and high bandwidth networks, and the evolution of satellite communications to networks of low earth orbiting satellites.
Computing and Software	The Computing and Software Area Team focuses on maintaining awareness of developments outside MITRE related to computer architecture and engineering, computer science, and software engineering.
Decision Support	This area focuses on cognitive-centered decision support applications and new methods and tools for developing effective systems that support decision-making. Emphasis is placed on decision-making in dynamically changing real-time environments (occurring in a day or less). Research in human decision-making to enable the development of better support systems for sponsors is covered in this area. Also covered is the demonstration of decision aids that advance the state of the art.
Electronics	Electronics investigates electronic component technologies, and their design and fabrication techniques.
Enterprise Architectures	Architecture development involves planning, designing, integrating, and managing complex systems of systems that can evolve to support changes in business needs and advances in software and information technologies. This area addresses the integration and interoperability of commercial components with custom-developed and current operational ("legacy") components.
Human Language	Human Language researches computer systems that understand and/or synthesize spoken and written human languages. Included here are speech processing, information extraction, handwriting recognition, machine translation, text summarization, and language generation.
Information Assurance	Information Assurance investigates security vulnerabilities in distributed information systems and develops architectures, systems and techniques for providing protection from attack, and exploitation.
Information Management	Information Management focuses on technologies and processes that enable the organization, creation, management, and use of information to satisfy the needs of diverse applications and users.
Intelligent Information	Intelligent Information Processing investigates technologies, tools, and

Processing	processes that support the discovery, processing, exploitation and dissemination of information, tools and knowledge. Intelligent agents are covered in this area.
Investment Strategies and Operational Analysis	The Investment Strategies (IS) technical area team is concerned with understanding the benefits and direction of planned and future technology investments by the government. Responsibilities include capturing information on trends in technology investments, understanding the challenges associated with investment decisions, and improving capabilities to support technology investment studies.
Modeling Simulation and Training	This area focuses on information technology to support training, and application of modeling and simulation. This includes advances in simulation infrastructure, interoperability architectures, and modeling paradigms. Additional emphasis is on the building simulations from reusable components.
Sensors and Environment	Sensors and Environment researches technologies to detect, monitor, and characterize the environment (terrain, weather, targets, etc.) to determine position within that environment (geo-position), and to manage, exploit and disseminate positional data (Geographic Information Systems). The use of radar, optical, sonic, and multi-spectral sensors is also covered.

Table 14-4: MITRE Technology Area Teams[136]

The Human Language[137] TAT is selected for illustration. The general page layout for the TAT is depicted in Figure 14-7 and includes: TAT theme, Team Members, and links to various resources and activities associated with the TAT. TAT reports and presentations provide a comprehensive description of the TAT's focus to include the core technologies, internal projects, sponsors, and prominent external organizations.

As a first strategy, TAT members may be assigned expertise descriptors based on the contents of stored reports and descriptions only. This first-order model is consistent with viewing the enterprise as a patchwork of non-overlapping activity spaces. Here, a TAT activity space is distinct from a project space even though a specific project may be linked to the TAT. For completeness, however, it may be useful to overview a more complicated second-order model that exploits structure between the TAT and TAT projects. This model reflects the fact that a TAT is an oversight or steering group for specific internal projects as particular domain. This more complex indexing model would then exploit the combined information and social space; that is, documents associated with TAT and projects as well as the combined membership.

---

[136] Abstracted from MITRE-internal descriptions

[137] The names of TAT members, specific project PI's, and references to sponsors are masked to ensure anonymity.
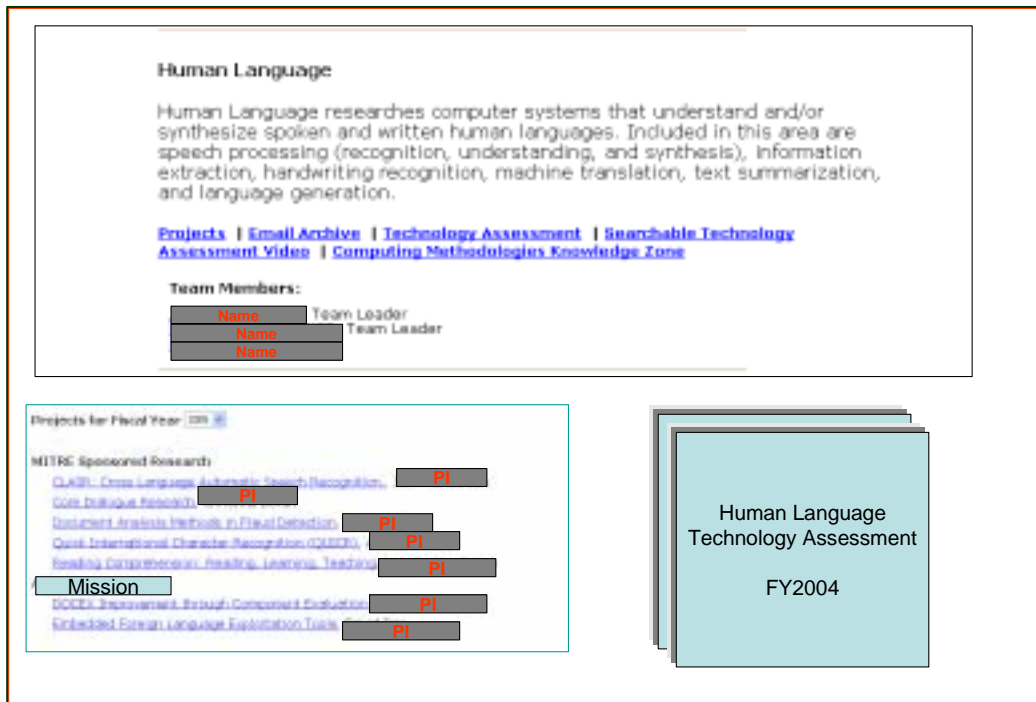
Figure 14-7:   Natural Language TAT Structure

From a collection perspective this model is fairly straightforward as project links are embedded in the TAT home page and are easily processed (lower portion of Figure 14-7). In this case the FY05 projects are displayed; however, there are links to earlier efforts as well. There may also be distinctions as to project type as an extension to the model. There are two classes of projects MITRE Sponsored Research projects and Mission-oriented projects which are collaborative projects with sponsor organizations. Partitioning internal research from collaborative work with the sponsor provides added flexibility for weighting projects according to maturity; under the assumption that sponsor projects are involved in technology transfer typically, while internal research is often exploratory work not expected to impact sponsors for several years.   As discussed, there are a number of indexing strategies possible depending on the tradeoffs between precision and computational cost.

The assignment of staff into specific TATs is a form of expertise signaling since TAT assignments require corporate approval. TAT products have corporate-wide visibility and are

subject to senior management approval before being published. TATs are anticipated to be a rich context for detecting "who knows what".

TATs may also be viewed as research project collectives. Depending on the TAT, research focus may vary considerably across projects or it may be cohesive and show significant work overlap in theme and in assigned staff. To gain a sense of how researchers are distributed across TATs a staff-overlap measure was computed for a number of the TATs. Figure 14-8 shows project overlap[138] represented as an *assignment overlap* graph; here, for the Sensors and Environment TAT. Note the relatively strong clustering involving five or so of the main projects (circled) and the other work that is on the periphery (at least in terms of the core staff). Assignment overlap provides insight as to the effective use of TATs for expert finding. For example, a researcher found relevant to a query on "IR Sensors" when juxtaposed with a TAT overlap map (as shown here) may be adjusted in rank depending whether they were in the "core" staff (shaded cluster), working on the periphery, or neither. This view could be further qualified by knowing whether projects were near completion or just beginning and representing new funding areas (possibly increasing relative weighting).

---
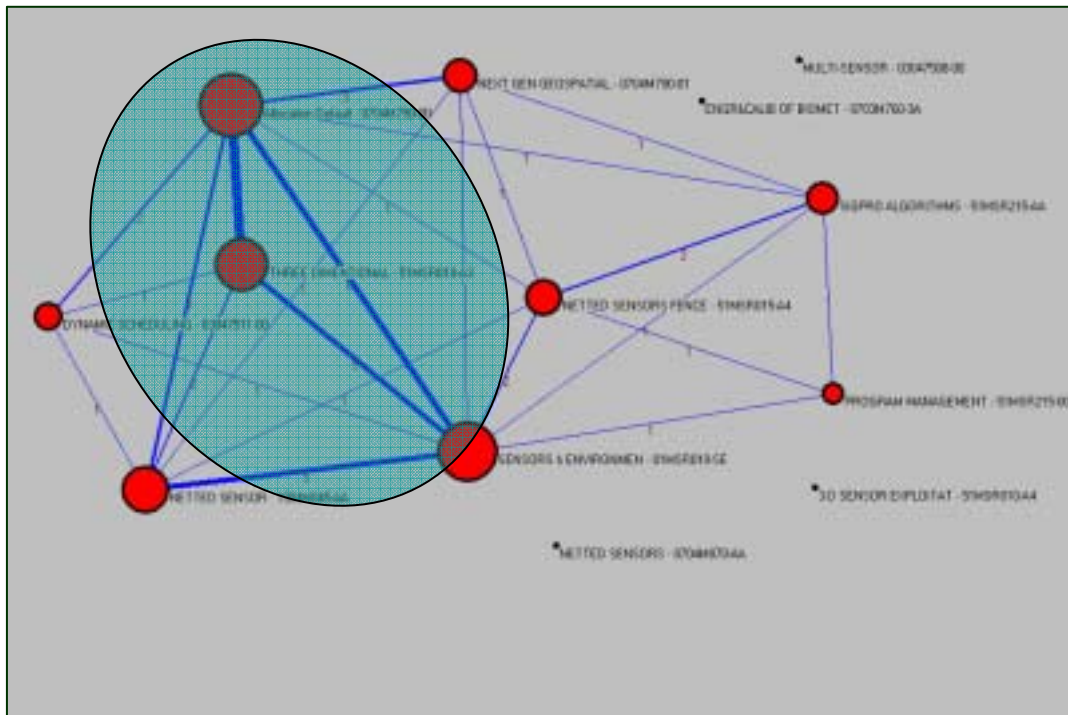
[138] Overlap was computed as the pairwise intersection between all projects in a TAT. Intersection is the number of people that co-work the project pair and is the edge weight connecting two project nodes in the TAT graph.

Figure 14-8: Staff Co-work Graph: Sensors and Environment TAT (core work shaded)

This notion of measuring the "crowd" affect (whether a person works the core research areas or is working as a relative isolate) is an aspect of social context which is accounted for in the expertise model (Chapter 6); for example by measuring activity space *density*.

14.2.6   HotLine

MITRE's Technical *Hotline* is an on-line service that lets users access resident experts in a number of technical areas. The *Hotline* service uses a peer-reviewed registration process for assigning experts into topic areas. Typically 3 to 5 experts are in each category. Users can email questions to *Hotline* experts using an online form that allows users to link questions to one of the 33 established expertise areas or to the "Other" category. The appropriate expert answers the question and archives the question and answer for searching and analysis of question trends.

Expertise categories are listed in Table 14-5, below and in Figure 14-9, the Knowledge Management area is shown as an example. Typically, as in Knowledge Management, there are several experts—while an entry for only one is actually shown. For each expert[139], contact information is provided along with the expert's picture and a description of any special skills the expert may wish to advertise (a form of signaling).

Acquisition Strategy	Antennas and Electromagnetics	Business Case, Investment and Technology Advise	Case Tools and Methods	CORBA	Cost and Schedule Analysis
Database Management	Digital Video	Electromagnetic and Nuclear Effects	Embedded Solutions Team	Enterprise Architecture	Informat-ion Warfare
Instructional Technology	Interoperability	Java	Knowledge Management	Linus	Mapping and Imagery
Mechanical Systems	Metrics and Measurement	Micro-electronics	Network Management	Operations Analysis	Perl
Quality of Service	Reliability	Risk Management	Software System Safety	Space Systems Analysis	Tactical Data Links
Unix Infrastructure	Web Applications Development	Windows NT	…Others		

[139] Expert identity is masked consistent with corporate policy.

Table 14-5: Hotline Expertise Areas
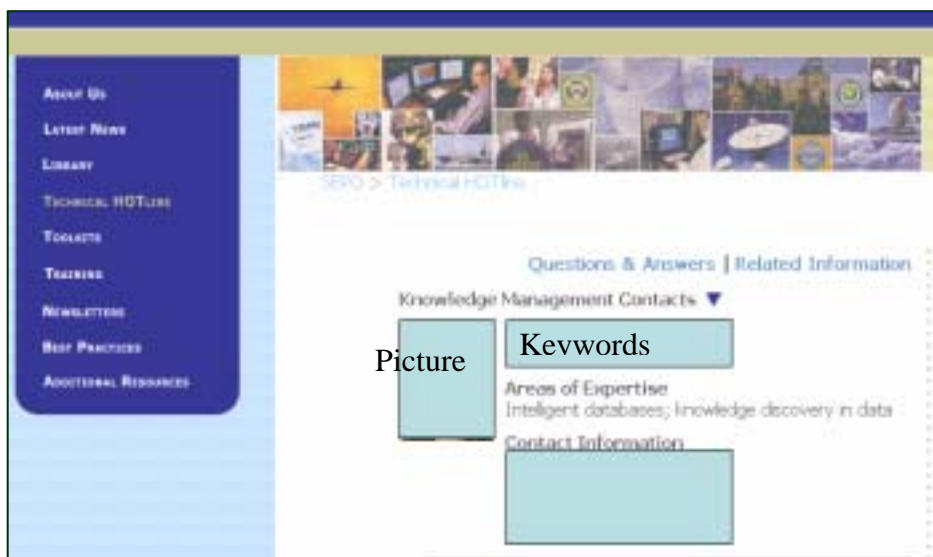
Figure 14-9:   Knowledge Management-- Anonymized

Most queries take less than an hour to answer, in which case there is no charge. If the response to a query requires more than an hour, the specialty area will provide an estimate of the effort required. Then the user can obtain project leader approval for the specialty area to charge labor hours to the user's project or overhead number before any work is done. Basically, this service in many ways tracks MITRE culture in terms of offering "free" advice unless extended effort is required. The online query form is shown in Figure 14-10. Note that while users must enter their employee identification number they may choose to remain "anonymous" in terms of the online Q&A archive.

The *Hotline* provides users with access to experts in a range of topics; however, in many cases users require access to experts that may have more familiarity with their actual problem or more specialized knowledge and they may prefer to work with experts that are members of their own organization. Because of this *Hotline* use is limited and most used to handle broad questions; in particular, questions regarding trends in industry or academia. From an expertise detection perspective, the *Hotline* is valuable to the extent that experts linked to expertise areas have been corporately "validated" by senior managers and peers. This registration provides some built-in

utility that should increase with scale-up in the question-answer archive. Having a baseline of experts may also provide context for identifying other (non-*Hotline*) experts that may have similar work profiles and that may perform like kinds of roles.

Figure 14-10: Online Question Form

14.2.7   MITRE Repository of Knowledge (MRoK)

MRoK is a knowledge management initiative focused on capturing knowledge directly from MITRE staff. Users post questions and answers to topic categories. MRoK has no formal registration in terms of an established cadre of experts; instead expertise is *attracted* to questions within one or more domains. The lack of formally recognized experts distinguishes MRoK from Hotline. The system, however, has similarities with ListServs in that domains are established and threaded discussions are possible. MRoK has some interesting features that exploit the question-answer formalism imposed. For example, a user can view posts by any individual employee that has contributed to MRoK making it possible to track individual interests and

expertise. Users can subscribe to any category in MRoK and in doing so, receive an email notification each time a new question is posted to the category.

Figure 14-11 presents a portion of the category hierarchy; users can review questions and answers by topic. A user can also review questions and answers associated with a particular person as shown in Figure 14-12. This has implication for extracting evidence of expertise based on questions answered and using that evidence in the *Expert Locator* prototype.

Figure 14-11: Partial MRoK Category Hierarchy

Figure 14-12: MRoK User-specific Questions

The search function, Figure 14-13, allows users to find questions and answers based on a topic (keyword) query or based on the contributor.

Figure 14-13: MRoK Search Function

Users can subscribe to a topic category and receive automatic email notification when questions or answers are posted, Figure 14-14.

Figure 14-14: Topic Subscription Form

With MRoK, expertise is made explicit through question-answer based interaction. The potential for gleaning expertise from MRoK is dependent on the quality and range of questions asked and answered. The level of interaction is largely driven by whether MRoK takes hold as an "everyday" tool or not.

14.2.8   Project Page

Projects reflect formal work, both internal and sponsor-funded. Projects typically have multiple tasks and each task has a membership list. Many staff members tend to work multiple projects and they often cluster around common work and technologies. Figure 5-14 shows staff (green nodes) with edges to multiple (yellow) project nodes (a one-to-many mapping).   The red ellipses suggest some work clustering; however transforming the 2-mode graph to show person clusters (based on co-work) or project clusters (based on shared staff) provides further insight.

The 2-mode affiliation graph in Figure 14-14 can be translated into a 1-mode co-membership graph, Figure 14-16. Here co-membership relates to personnel sharing across project pairs.

Therefore, a link between two projects relates to the number of people working both projects. To better reflect which projects have the most joint work, each project node is sized according to the total shared labor it has with other projects. Isolated projects will have few links to other projects and few members—there nodes will be small. A first-order project space model may reflect project size (membership) or sponsor domain, while more complicated models may factor the project dependencies such as the links between projects.
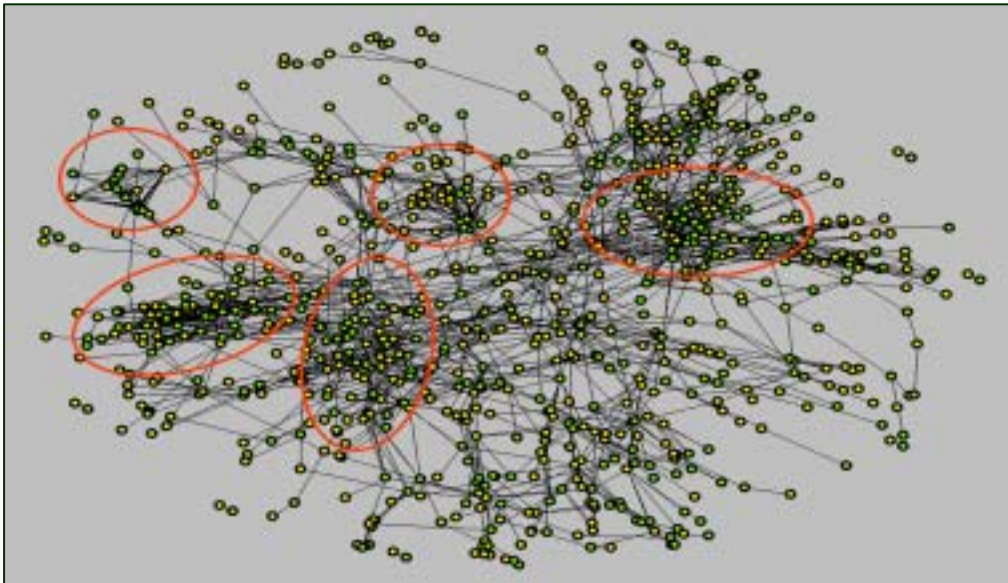
Figure 14-15: Business Area Project Graph: Yellow Project Nodes and Green Person Nodes
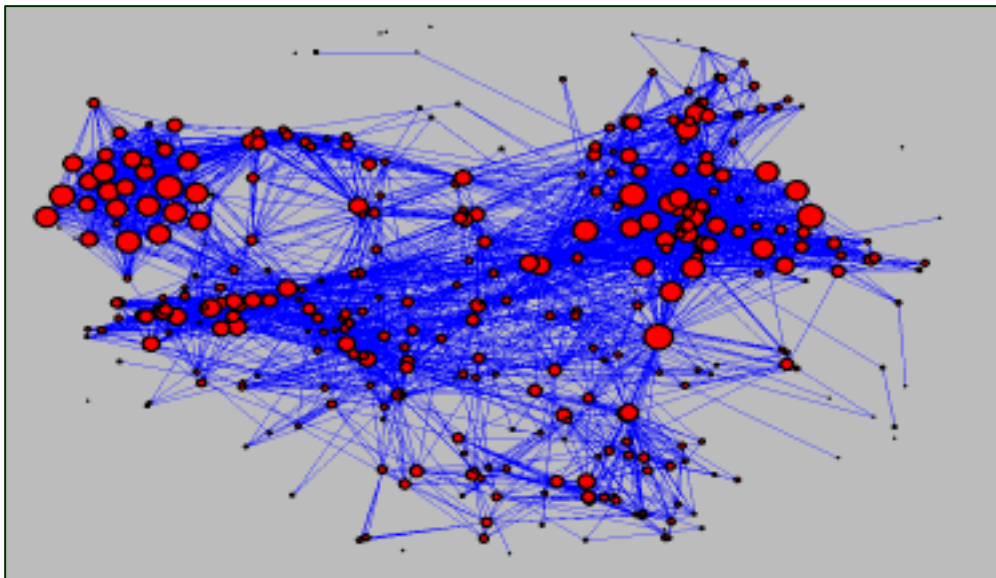
Figure 14-16: 1-mode Co-membership Project Graph

Finally, the 2-mode graph can be translated to another 1-mode co-work graph showing ties between staff members, Figure 14-17. Two staff members are linked if they have joint (i.e., overlapping) work. This is very similar to authorship graphs that depict the collaboration between multiple researchers. Note also that nodes are sized based on their level of "collaboration"; that is their total co-work with others. This project space view suggests that some staff may play broad (multi-project) roles while others are isolated to one or a few efforts. This has particular significance if all projects are relevant to a particular expertise area.
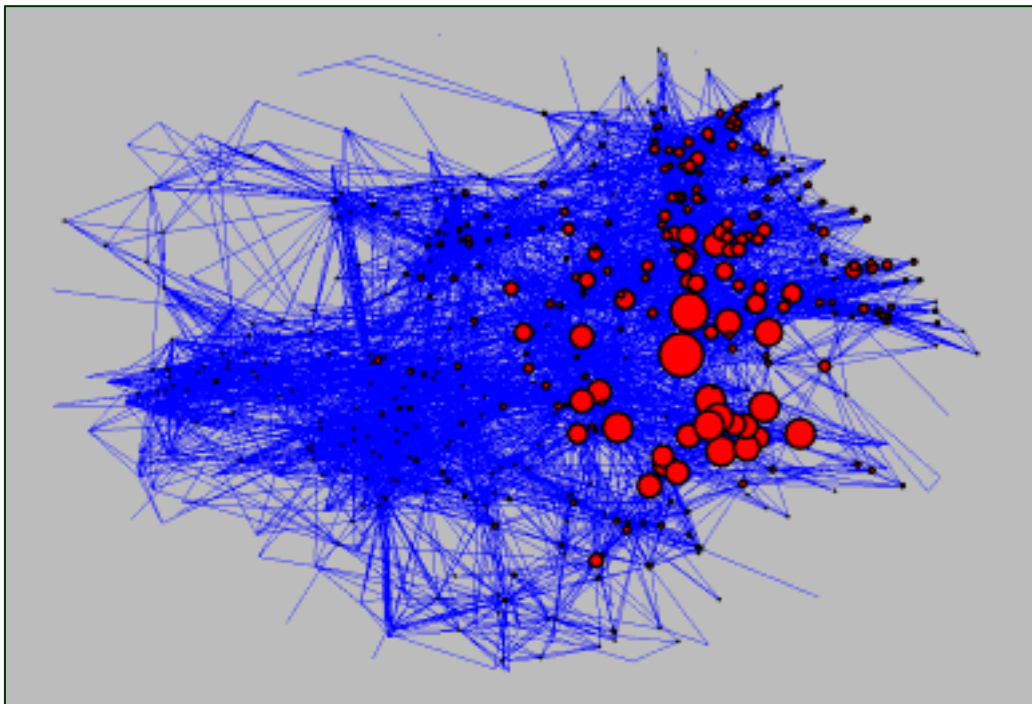
Figure 14-17: 1-mode Co-work Personnel Graph (Staff Names Masked Out)

These views of project landscapes reflect the complex nature of work and provide insight as to how expertise might be distributed across actual work areas. In particular, analysis of the "whole" multi-project graph provides insights into which projects and people may be the best connected workers. The best connected experts may also be important to identify. This is addressed further in Chapters 6 and 7. The focus shifts to project structure and data organization.

Projects consist of tasks, task members, labor charges, events, and artifacts. Project data are split between a Project Page which includes standard project metadata (such as task membership, labor charges, owning organization, and sponsor affiliation) and the Project Share Folder which

contains task artifacts or documents. Project metadata may be used to tailor task membership lists and this is often done by managers to better assess which staff are most actively involved in the project. For example, the task leaders can be identified and task members ranked by labor usage. Knowing the distribution of work across a task allows membership "filters" to be set up that can constrain the task list to the highest ranking members. This may be especially useful when reducing large task membership lists that are largely made up of staff that charge very little actual time; for example, they may have very limited consulting or management roles. Documents are linked to tasks when task members publish items to the appropriate Project Share Folder. Next, Project Page and Project Share Folder structure are discussed.

A representative project page is shown in Figure 14-18. Each page includes the task name, a short description (label), the parent project (as most projects have multiple tasks), task leader, and period of performance. This is followed by a list of task members to include their home department and level of effort (percentage of total task labor used).   The links on the left side or the page include Project Share which contains all documents archived to the task.

Figure 14-18: Project Page

Key project data such as membership lists and labor usage was obtained by parsing online project pages using a Perl script. This type of "screen scrapping" is necessary since direct access to the underlying project data (a corporate database) is prohibited based on access policy restrictions imposed. Using Project Share folders it is possible to associate project documents with project team members and to combine labor usage and technical level as a basis for gleaning who may be providing key technical contributions.

14.2.9   Project Share

The Project Share Folder system is a Web-based environment for knowledge sharing and reuse. Project Share allows MITRE users to publish and share project-related documents and files. Access to documents and files shared is available by browsing folder hierarchies, or by searching. Documents and files can be published, browsed, and viewed in their native file format versions. Figure 14-19, provides a typical Project Share view. Key to this discussion is the document list that is linked to the project task.
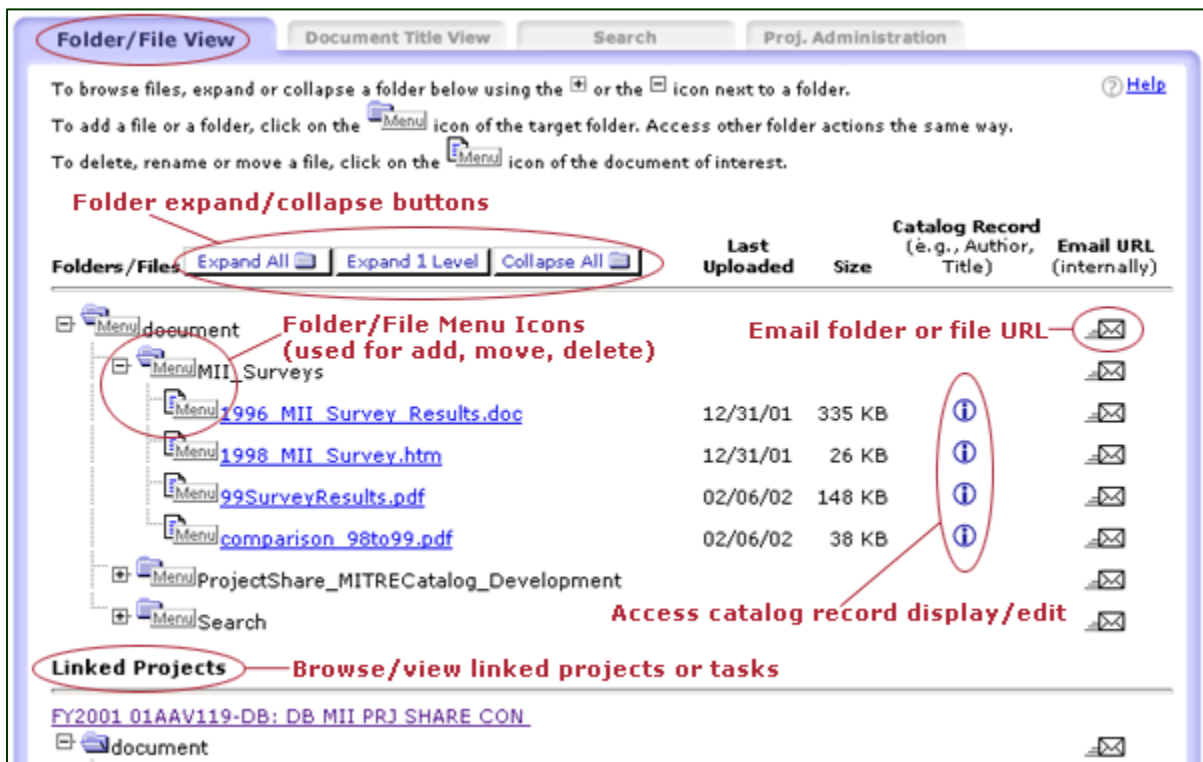
Figure 14-19: Project Share Space Mapped to Exploitable Data Structures

14.2.10   SourceForge

There are a number of online software development environments and scientific laboratories that may be useful in identifying new developments, ongoing experiments, key technologies, and areas of specialization. For example, there is now a SourceForge[140] server (iSF) internal to MITRE. It provides developers with access to a wide range of tools including bug tracking, task management, code versioning (CVS), mailing lists, forums, and project web pages. These unified tools provide developers and project managers with the necessary resources to focus attention on project development instead of project management. The developer has control over his/her own project; controlling who can change information, what services the project uses, etc. It is a pilot service for MITRE internal use only. There are currently over 200 hosted projects with 544 registered users. A description of selected functions is provided in Table 14-6, below.

Source Forge provides a potentially rich environment in which to glean expertise. iSF can provide evidence of software development or application expertise and can be used to identify development teams and link teams to sponsor programs.   Consistent with the notion of applying expert finding algorithms customized to a particular activity space, it is possible to develop multiple methods here for measuring expertise; for example, based on individual programming language skills. Unfortunately, iSF arrives late with regard to this research and therefore it could not be considered for integration into the prototype development. Even cursory assessment of the iSF pilot program was problematic given that it took a number of months before a significant number of projects were registered. However, iSF will be evaluated for future integration into *Expert Locator*. At this time, however, these data are not used in the *Expert Locator* model.

---

[140] http://sourceforge.net/ Accessed on October 15 2005.

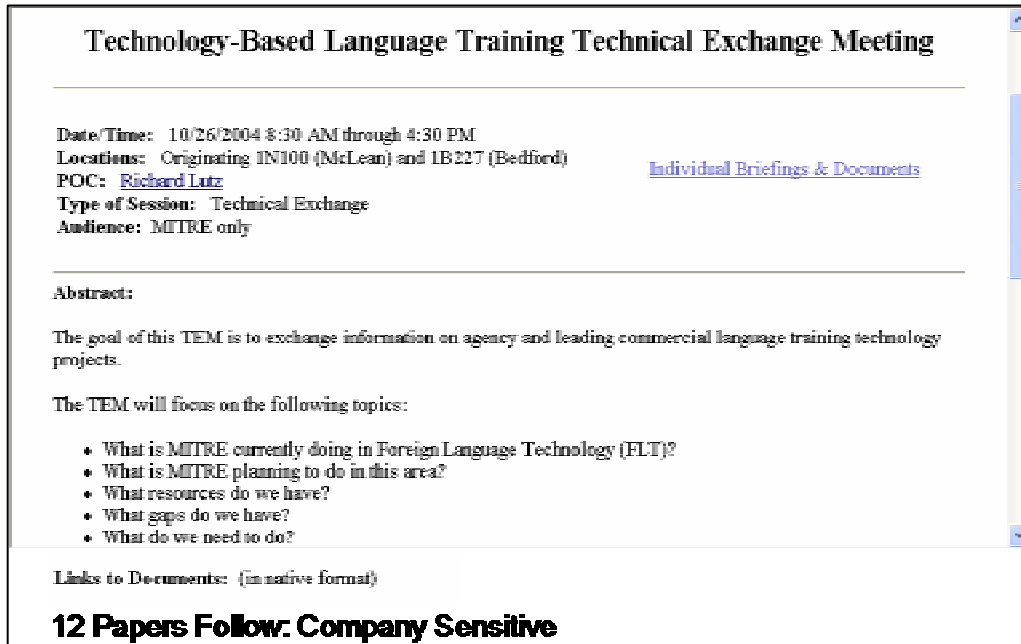| Functionality	Description
Basic Project Information	Administration, Developers, Project Type, and other organizational information tied to the Project
Home Page	Web pages for the project
Source Code Repository	Storage for project source code. Users can browse source code repository. Users can access information on individual files, to include change logs, source code, and versions.
File Releases	Formal releases of your software. Users can make snapshots of source code for downloading.
Mailing Lists	Projects can be linked to MITRE mailing list(s) allowing anyone browsing a project to send an email to team developers.
News	News items can be submitted to a project and displayed on a summary information page. Way to submit news about your project.
Forums	Discussion Forums can be set up for the project and each forum can be monitored for new postings.
Trackers	There are four default trackers for every project: Bugs, Feature Requests, Support Requests, and Patches. Users can create new trackers to track additional items.
Document Management	Documents can be published and linked to the project.
Task Management	The Task Management System allows users to manage project tasks to include affixing start and end dates, monitoring labor usage, and alerting developers as to changes in tasking.
Surveys	Users can create surveys and gather user feedback for a project.

Table 14-6:   Selected Source Forge Capabilities

14.2.11   Technical Exchange Meetings

Meetings are critical to coordinating activities and sharing information. While informal, ad hoc meetings involving small groups may dominate in some environments, formal meetings are critical business activities. Formal meetings often link to descriptive information as to meeting purpose, topic, attendees, and results. Where meeting results are archived, a number of options exist for identifying expertise, groups of related experts, and related artifacts such as briefings or technical papers.

Technical Exchange Meetings (TEMs) are internal meetings held by MITRE employees for MITRE employees. Generally, technology experts or business stakeholders organize a TEM to meet a compelling technology or business issue. TEM organizers are usually responsible for all aspects of the TEM; from specifying the theme, to putting out the call for presentations, to setting the location and time. Figure 14-20 is a typical example of a TEM announcement.

Figure 14-20:   TEM Announcement

TEMs focus on key issues and technologies that are currently or projected to impact MITRE sponsors. TEMs may be quite useful for identifying expertise. Each TEM is focused on a theme. Topics vary and include content management, Biocomputing, cross-boundary information sharing, secure mobile wireless devices amongst others. A TEM typically has multiple sessions, and presenters. Each TEM is archived to include attendance lists, briefings and papers, and summary findings. A TEM includes a brief description consisting of the TEM theme, target audience, and registration requirements (if any). TEMs are modeled on the workshop format in which topics are well focused, involve some formal agenda with attendee participation incorporated within, and have a formal registration process to control attendance. TEMs generally receive corporate support to include access to needed facilities, and arrangements for

video and textual data capture and archival. The byproduct of a TEM is often a collection of briefings, papers, and a summary result. While not intended to ferret out areas of expertise, TEM outputs are potentially useful sources of expertise indicators.   In particular, there are several bases for attributing documents to authors and for identifying the TEM organizers. While TEM documents may also be accessible through the corporate search engine associating the documents with the TEM provides a basis for adding additional context to individual items.

Many TEMs recur annually and tracking TEMs over time is important for assessing evolutionary changes in technologies or operational problems areas. This may also have implications for tracking the evolution of an expertise network; changes in membership and work relationships. For example, Figure 14-21 provides an affiliation graph that depicts attendance patterns across four TEMs addressing enterprise architecture.   There are 4 nodes marked as "TEMs" signifying the meeting; each TEM label provides a description of the main TEM theme. The rest of the nodes are unlabeled (anonymized) MITRE staff members. The staff nodes are sized and colored so that the larger the node the more TEMs that person attended; the maximum is 4. Clearly, there is a core group in the graph center that attended multiple TEMs; however most attendees participated in only one TEM. Since each TEM focused on a particular sub-theme under enterprise architectures there seems to be at least four sub-populations in the overall community. Identifying expertise sub-areas as well as expertise networks or communities is important in terms of characterizing expertise across the enterprise and changes in expertise over time. Note the enterprise architecture TEMs occurred over a four year period.
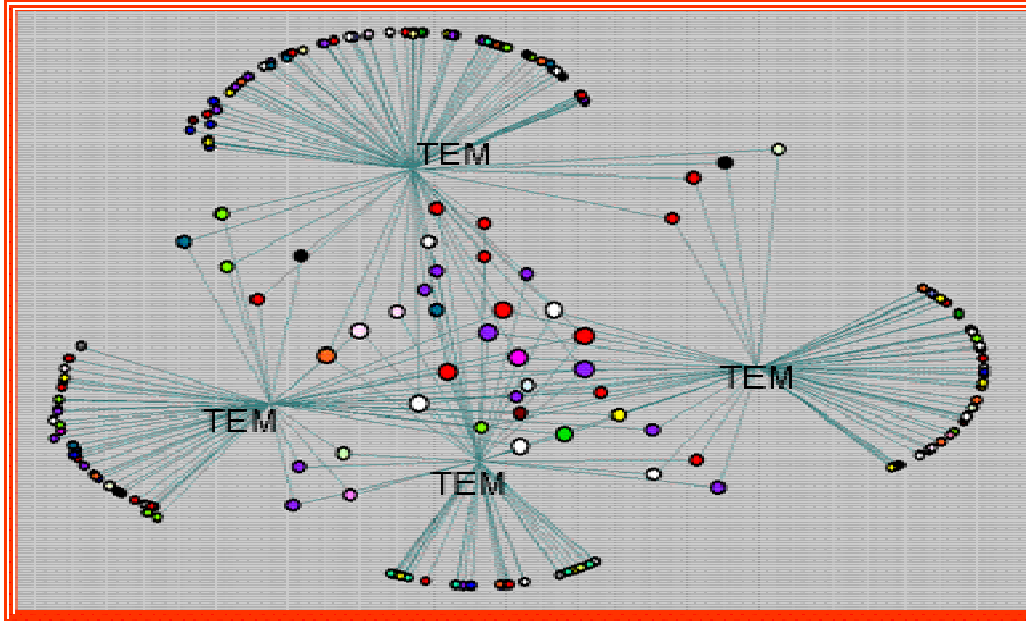
Figure 14-21: Enterprise Architecture TEMS across four years

The Technical Exchange (TEx) is an application that provides a centralized location for all MITRE employees to submit and view information about Technical Exchange Meetings, Conferences, Seminars, or Symposia; or post Requests for Papers for these types of meetings. Briefings, papers, and other documents associated with these meetings are available through this collection. Currently, TEM materials are searchable using the enterprise search engine and this makes them easily exploitable by an expert finder system; however they are not directly attributable to the supporting TEM as at the time this research was conducted TEMs were not organized under the TEx system nor were TEM documents organized in any recognizable directory. This made document-TEM attribution problematic and precluded the use of TEMs in the initial prototype (see Chapter 7).

14.2.12 ListServs

A ListServ may be viewed as a mechanism for forming communities; a sort of *communityware* that supports self-organization around selected business or technology issues. ListServs are essentially open forums for dialogue on various topics. In many cases, dialogue is constrained to the core topics or issues associated with ListServ; although this can vary widely and users are often *OT* (off topic). For example, an Analysis Tools ListServ may be quite diverse in terms of

273

the types of tools, the enabling technologies, tool use and evaluation. As such it is problematic to associate all members with the main focus of a particular ListServ; a finer grain view is needed; individual postings

A posting is essentially an email sent to ListServ membership. This simple mechanism supports open dialogue amongst a focused audience, and makes key issues visible in ways that go beyond standard e-mail. The body of the posting may be analyzed topically and also to reveal threaded conversations. Header information may be used to identify the ListServ, sender, date/time posting and other information depending on structure displayed. Beyond the simple post, users can easily obtain membership information and this makes it possible to track users as to posting behavior over time, as well as identify lurkers who do not post but may be reading posts only (a type of free-rider effect).

Overall, low-level posting behavior provides for a rich social context in which to ferret out expertise. This includes finding experts that may be thought leaders within their community (possibly on issues not related to their expertise). As an example, an expert, say on support vector machines, may also be a thought leader on data privacy within the same ListServ; interacting with a significant number of members. While in this example, being a thought leader may not influence an expertise rating directly; it could be used as a secondary basis for selecting an expert to contact; especially when looking for experts that have broad issue knowledge.

While individual posting behavior (e.g., specific topics posted on, numbers of posts, participation in threaded discussions) may be useful for inferring which ListServ members may have certain knowledge or expertise, posting behavior across multiple ListServs may be used to assess global (community-of-communities) structure. Linking individual ListServs based on co-membership can be used to generate a *fitness* landscape that can suggest which ListServs are most useful in discriminating amongst members. This parallels the standard use of discrimination measures in information retrieval. This global analysis may also be used as a separate basis by a user for selecting experts. For example, if experts are modeled as having high degrees of specialization then participation in multiple ListServs may suggest they have generalist qualities. A

specialization[141] factor of this type may not be the primary factor in ranking experts or selecting them for contact but it could enhance retrieval list output. For example, the (complex) graph shown in Figure 14-22 shows 2137 ListServ members from 267 ListServs where they are ranked from low specialization (bottom of graph) to high at the top. This inverted pyramid places those with highest specificity at the top. The lowest rated staff member may have broad based knowledge and be especially useful in providing referrals. As it turns out the actual person (here anonymized as #1512) is actually a coordinator working across multiple research programs and has wide-spread operational duties as well. Again, all this serves to illustrate that ListServs as communities are rich social contexts that can be used to enhance retrieval results.
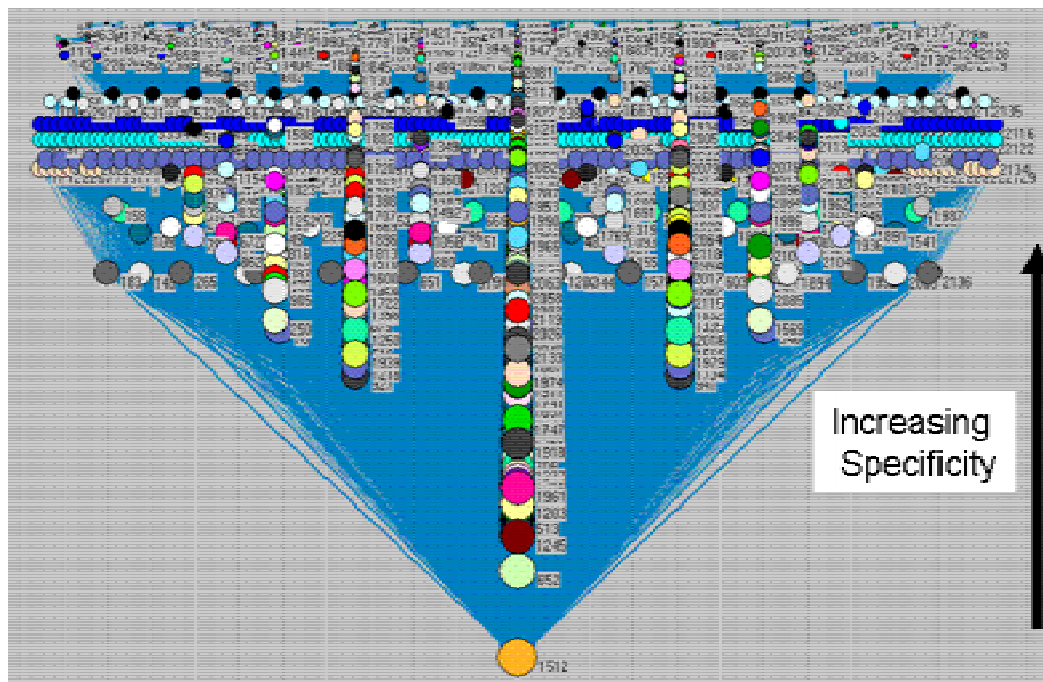

Figure 14-22: Degree of Specificity Used to Illustrate Social Context in ListServs

ListServs are potentially rich community spaces that may capture wide-ranging formal and informal work. Specific Lists may focus on particular technology areas or sponsor domains and provide a rich context for gleaning expertise; as represented in Table 14-7.

---

[141] Here, specialization is essentially a coverage measure used to assess the actor's diffusion across ListServs.

Selected Technology-centric ListServs
Grid Computing Discussions
Web Services
Airport Metrics
Infectious Disease Surveillance
Autonomic Computing Discussion within MITRE
Foreign Language Exploitation
Engineering Complex Systems of Systems
Data Mining and Knowledge Discovery
Topics Relating to Database Technologies
Ethnography and Qualitative Research Discussion Group

Table 14-7: Selected Technology-centric ListServs

From an *Expertise Relevance* perspective, ListServs are highly applicable to gleaning expertise based on the large number of discussion forums supported. In addition, ListServs have significant activity based on average posting frequency (*Artifacts*), and overlapping memberships promoting cross-disciplinary dialogue and information diffusion (*Social Presence*). For example, the top 1000+ ListServs ranked by membership size, Figure 14-23, suggests that a number of ListServs are large enough to support rich interaction; the average size is 37 members with a median size of 15. There are more than 500 Lists with 15 or more members. As shown in the Figure 14-23 insert, there is roughly linear growth in postings over time; here approximately 362000 postings were generated across 2000 ListServs over a six month period ending in December, 2005. This suggests ListServs are both artifact and socially rich spaces that are used heavily to address a wide range of issues and topics not easily made visible through normal project work.
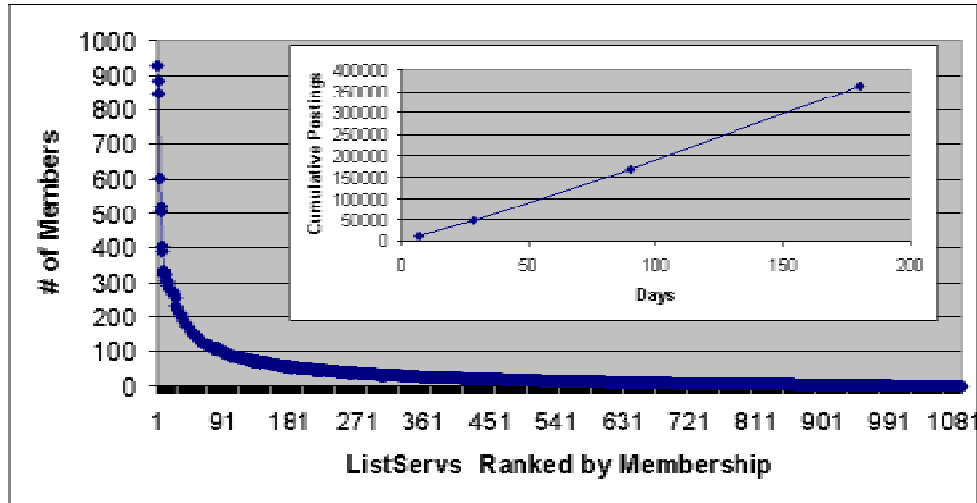
Figure 14-23: ListServs Ranked by Membership Count with Postings Statistics as an Insert

ListServs are generally not *stovepiped*; topic overlap across Lists is not atypical and this provides added basis for dispersed experts to weigh in on what might otherwise be narrowly channeled discussions. Essentially, cross-posting may reflect interlocking Lists and the potential for information diffusion. However, a more general view of cross-posting is provided in a series of *snapshots* in Figure 14-24. Here, information diffusion is characterized indirectly from the perspective of community linkages; that is, overlapping membership. This does not constitute actual traffic analysis, but does suggest the potential for capturing rich interaction across diverse memberships.

Starting at the top left, actors are linked to ListServ[142] based on membership; there is a large connected component that subsumes most of the ListServ population. This suggests that most ListServs are connected through co-membership at various degrees; some more closely connected than others. There are several "islands" that are isolated from the main core; however, these are in very narrow specialty areas. The figure, top right, highlights Lists according to communication volume (number of postings), and, as expected, a significant proportion of the overall traffic can be relegated to a few dozen or so Lists; this power law distribution is typical of most social networks.

[142] Actors are represented by circles and directed arcs point from actor nodes to affiliated ListServs.
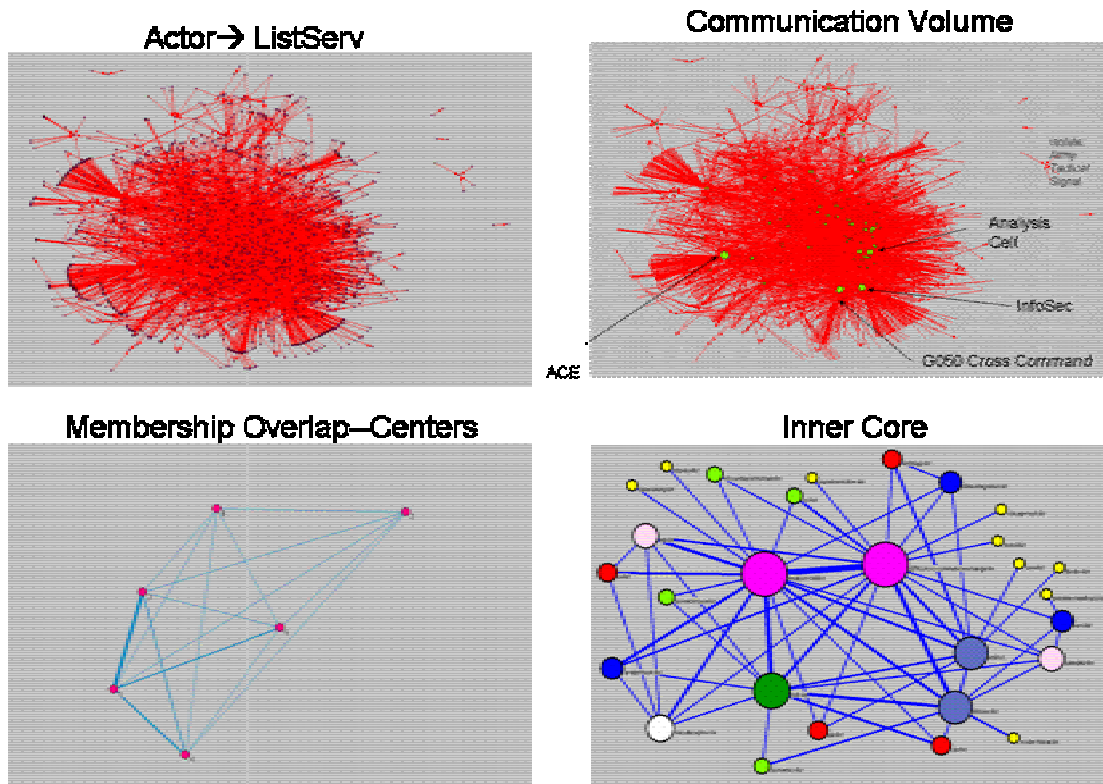
Figure 14-24: Various Views of ListServ Interconnectivity

The two-mode (Actor-ListServ) graph discussed so far is transformed into a 1-mode (ListServ-ListServ) connection graph as shown in the Figure, bottom right. This graph reflects overlapping membership across all ListServ pairs. However, here, the "inner" core is shown, that is, weak edges are cut leaving only the most densely connected Lists. This further confirms the interrelationship between certain Lists, and reinforces the notion that there is the potential for information diffusion across Lists. This overlap is observed at the corporate center level as well; as shown in Figure 14-24, bottom left. Of the six centers shown, there is relatively higher overlap between two centers (based on edge thickness); two others have "secondary" coupling to the strongly tied centers and the remaining two are weakly tied to all others.

While nearly all ListServs are quite active, there is a central core, that shows significant overlap in membership and this may be tied to business units that perform joint work. This suggests that ListServs are likely to be central to communicating key business issues, especially regarding discourse on enabling technologies and overlapping problem areas. More formal business

functions are likely to be handled using other mediums that better guarantee privacy or support legal reporting requirements. This would include standard business correspondence, email, and face-to-face interaction. For the most part, ListServs form multiple overlapping communities of interest that may have significant value in the analysis of "who knows what".

14.2.13 Community Share

Community Share is a pilot project using a community-based document management product, Sharepoint[143], to address MITRE's requirements for team support. Sharepoint is a community- or team- based collaboration platform that provides a common web space for working on shared documents, posting events and announcements, posting links to web sites, having threaded discussions, and tracking action items or agenda items. A representative "homepage" is shown in Figure 14-25.   As a content and social space it has significant potential for supporting expert finding; for example with regard to the following three perspectives:

- It provides a number of capabilities for supporting communities of practice; to include content management and community interaction
- It effectively competes with existing MII services and resources. This has implications for how *Expertise Locator* may work in the future; to include replacing some current activity spaces with Sharepoint.

Figure 14-25: Sharepoint Community Home Page (Human Centered Design)

---

[143] http://www.microsoft.com/sharepoint/ Accessed on 18 October 2005.