MTR 06B000040

MITRE TECHNICAL REPORT

Expert Finding Systems

September 2006

Mark T. Maybury

MITRE

Center for Integrated Intelligence Systems
Bedford, Massachusetts

CIIS
Center for Integrated Intelligence Systems

MTR 06B000018

MITRE TECHNICAL REPORT

Expert Finding Systems

June 2006

Mark T. Maybury

Sponsor:	MITRE Corporation	**Contract No.:**	
Dept. No.:	G060	**Project No.:**	05AAV010-JR

Approved for public release; distribution unlimited
Case No. 06-115.

MITRE
Center for Integrated Intelligence Systems
Bedford, Massachusetts

MITRE Department Approval: Dr. Mark Maybury
Executive Director
Information Technology Division

MITRE Project Approval: Dr. Lisa Costa
Chief Engineer
Commands, Technology, and
Intelligence Integration

Abstract

The ability to rapidly discover individual experts, communities of expertise, or knowledge artifacts created by experts is an essential element of organizational effectiveness. This report outlines requirements for, challenges to, current state of the art in, and commercially available services and solutions for expert finding. The report provides a matrix of commercial off the shelf expert finding tools, characterizes their capabilities, and provides best practices to ensure a successful deployment.

KEYWORDS: Expert finding, expert location, expertise, commercial off the shelf (COTS), knowledge management.

Acknowledgments

We thank Carol Elliot for her help in collecting vendor information and Paula MacDonald for setting up interviews with vendors. We also thank each of the expert finding vendors for their time and energy in clarifying their product offerings, histories, and futures. We thank Dave Poutre for his mission advocacy of this effort and Dr. Lisa Costa for her project support of this study. Appreciation to Pam Davis who proofed the final report.

Chris Elsaesser had the original idea for an expert finder for MITRE and David Mattox, Inderjeet Mani, and David House helped create MITRE’s prototype Expert Finder. Manu Konchady and Raymond D'Amore investigated expert network discovery at MITRE. MITRE’s operational expert finder was created by Sean Cooper and Robert Joachim.

Table of Contents

Figures

Tables

Executive Summary

Expert Finding Systems (EFS), also called Expertise Location Systems (ELS) enable users to discover subject matter experts in order to hire or acquire their knowledge. EFS can make organizations more efficient and effective by rapidly locating individuals or communities of expertise to accelerate research and development, enable rapid formation of operational or proposal teams, or support formation of cross disciplinary teams to respond to new market threats and opportunities. EFS can also be used to assess enterprise skill sets, enabling the identification of skill atrophy, the discovery of new and emerging skill areas, or the prediction of the effects of skill loss (e.g., as a result of attrition or retirement) or gain (e.g., as a result of a merger or acquisition).

EFS need to support a number of key requirements including the ability to:

- *Identify* experts via self-nomination and/or automated analysis of expert communications, publications, and activities.
- *Classify* the type and level of expertise of individuals and communities.
- *Validate* the breadth and depth of expertise of an individual.
- *Recommend* experts including the ability to *rank* order experts on multiple dimensions including skills, experience, certification and reputation.

Expert finding is a difficult task because experts and their skills and knowledge are rare, expensive, (unevenly) distributed, difficult to qualify, continuously changing, varying in level, and often culturally isolated and oversubscribed. To complicate this, expert seekers typically have poorly articulated requirements, are ignorant of expert’s past performance, and are not fully enabled to judge a good expert from a bad one. Finally, their complex problems often require the combined wisdom of multiple experts.

Expert finding has been an active research area for several years. For the first time, the Text Retrieval and Evaluation Conference (TREC) Enterprise track evaluated 9 research systems in a task to find World Wide Web Consortia experts on 50 topics. Using over 300,000 documents retrieved from the web (*.w3.org) to automatically build expert profiles, the best system achieved a Mean Average Precision (MAP) of 27.5%.

Several COTS tools have become available that automate the discovery of experts. These include TACIT ActiveNet™, AskMe, Autonomy IDOL K2, Endeca, Recommind, Triviumsoft’s SEE-K, and Entopia Expertise Location. Each product is assessed in detail based on literature analysis, vendor interactions, and demonstrations. Products are characterized and contrasted in terms of the sources processed, the kinds of processing performed, types of searching supported, kind of results presented, properties of the system (e.g., interoperability, privacy), type and size of deployments, and cost. A summary of the results are shown in Table 1. These systems been applied to most industries including pharmaceuticals, healthcare, financial services, professional services, information technology, aerospace, manufacturing, media/broadcasting, retail, state and local government, defense and intelligence, and academia. Successful deployments of EFSs require executive championship, involved users, user/culture centered design, clear purpose, realistic goals, measured usage and benefit, simplicity, ease of use, incremental deployment, appropriate privacy, incentives for use, and effective marketing, communication, and training. While financial return on investment has been difficult to characterize, multiple organizations report cost savings, time savings, and new business opportunities.

1. Introduction

The ability to rapidly discover individuals or communities of expertise is an essential element of organizational effectiveness. Experts can answer questions, point to definitive sources or specialists, or perform needed functions requiring special knowledge, skill or experience. This report details a trade study that explored the requirements for, challenges to, current state of the art in, and commercially available services and solutions to enable expert finding in an enterprise. We conclude by providing a matrix of commercial solutions and characterize their capabilities, similarities and differences, and availability.

Expert Finding Systems (EFS), also called Expertise Location Systems (ELS) enable users to discover subject matter experts in order to hire or acquire their knowledge.

1.1. Objective and Goals

The primary purposes of this study are to:

- Describe the state-of-the-art EFSs both now and likely in the next 18 months.
- Identify features provided out-of-the-box and via customization or integration with other systems.
- Assess other operational, management, and technical characteristics (e.g., availability, flexibility, robustness) of EFSs.
- Identify the technical, social, management (e.g., vendor viability, market share), etc. issues that sponsors need to consider to successfully implement an EFS.
- Identify critical factors and metrics such as return on investment (ROI), development and sustainment costs, integration with other systems, etc.

1.2. Study Methodology and Scope

The study initiated with a literature search of expert finding research in academia, industry, and government. This is summarized in the related research section. Subsequently, expert finding solutions that were commercially available were identified from a broad range of sources include internet searches, discussions with search, collaboration, and knowledge management vendors, review of deployments at vertical business areas (e.g., pharmaceutical, finance, government, aerospace, and energy). Following analysis of product literature and web sites, direct interviews and in some cases demonstrations of products were conducted. During this process many on-line expert finding services were identified (e.g., for finding expert witnesses, medical expert recommenders), however, the scope of this study focused on commercial of-the-shelf systems that could be leveraged for rapid deployment in a defense or intelligence enterprise.

1.3. Report Structure

The remainder of the report is organized into the following areas of expert finding:

- *Requirements:* What are the user and organizational needs driving EFSs?
- *Challenges:* Why is expert finding a difficult task?
- *Previous Research*: What has already been learned about EFSs?
- *Performance*: How well do EFSs work?
- *Commercially Available Tools*: What COTS tools are available to automate the discovery of experts?
- *Lessons Learned*: What lessons have been learned about best practices to ensure a successful EFS?

Readers interested in learning about the rationale for and challenge of EFSs can read the first few sections. Those interested in current state of research and about MITRE’s Expert Finding system can read the research and performance sections. Those interested in available commercial tools can read that section. Finally, those interested in lessons learned and best practices from deployments can read the final section.

2. Requirements

Distribution of staff, decreasing project size, and cost/time pressure are driving a need to leverage enterprise expertise by quickly discovering who knows what and forming expert teams (Fenn 1999). Those in need typically have little or no means of finding experts other than by recommendation. At the other end, busy experts do not have time to maintain adequate descriptions of their continuously changing specialized skills. Past experience with "skills" databases indicates that they are difficult to maintain, quickly outdated, and reflect self-reporting biases. What is required is an ability to support the following functions with respect to experts:

- *Identify*: Cull through explicit (self) nominations and/or large collections of artifacts (email, instant messages, documents, briefings) created by individuals to implicitly determine candidate experts in a given topic.
- *Classify*: Assess multiple sources of evidence to characterize the type and level of expertise of individuals. Analyze competencies and relationships among experts to determine communities of interest and communities of practice.
- *Validate*: Assess the breadth and depth of expertise of an individual to verify their expertise level. Expertise qualification can be done by human assessment, marshalling evidence (e.g., qualifications, resume, publications), or automated user feedback mechanisms (e.g., positive and negative feedback from interactions to establish reputation.).
- *Rank*: Produce a rank order of experts on particular dimensions (e.g., years of experience, type of experience, certifications, publications, etc.).
- *Recommend*: Given a particular information need and importance criteria (e.g., breadth vs. depth, types of experience), return a rank order list of experts or expert communities that are most relevant to the need.

Yiman-Seid and Kobsa (2003) outline some of the reasons for seeking an expert (as opposed to a document or searching) including the need to access non-document information, the need for expert dialogue to identify or specify a problem, the need to use the expert to filter large amounts of information, the need to interpret or contextualize information (including perhaps assessment or recommendation), and a social need, that is a preference for human interaction. In other cases the seeker wants an expert to perform some task, e.g., to become an employee or contractor, a team or committee member, or to act as a speaker, teacher, or expert witness/interviewee.

Challenge Statement: To create tools and methods to enable rapid, accurate, inexpensive and privacy/security sensitive discovery of experts and expert networks and the knowledge they contain.

3. Challenges

Expert finding is challenging for many reasons including:

- The volume of communication/publication is no indication of expertise.
- The first expert you find may not be best one.
- Certain topics engender more opinion than facts and so finding the true expert can be difficult.
- There generally is a lack of access to information about past performance of experts.
- New employees don’t know about informal social networks hence cannot exploit these to find experts.
- Privacy concerns may limit the degree to which measurements of expert performance is shareable.
- Expertise is not distributed evenly and strengths of associations among experts vary significantly.
- There are no standards specifying the criteria and/or qualifications necessary for particular levels of expertise.
- True expertise is rare and expensive. Often access is controlled, either informally or formally, either by the expert themselves or their management.
- Expertise continuously changes and requires awareness of this dynamic.
- Solutions to complex problems often require either communities of experts or diverse ranges of expertise that need to be brought together to solve the complex problems.
- Engineers in one classic study spent 16% of their time communicating with experts – but communication was impeded by geographic, time difference, and cultural barriers.

In summary, expert finding is a complex and difficult task.

4. Previous Research

Semi-automated methods for discovering and assessing expertise have been investigated for at least 15 years. For example, in the Dataware II Knowledge Directory[1], experts can self-nominate and subsequently be discovered through directory search, however, this manual process is expensive to maintain and becomes quickly out of date.

Swartz and Wood (1993) investigated email flow and not content analysis to identify (but not rank) "distinguished" people. The ContactFinder (Krulwich and Burkey 1996) system analyzes addresses and content of bulletin board messages to determine who best can answer questions about a particular topic.

Campbell et al (2003) at IBM compared a content based approach that looks only at email content and a graph based method that looks at social networks from email communications and found in the email expertise extraction (e^3) system the latter outperformed the former in two different organizations (a research (OrgA) and a software development one (OrgB)). They analyzed 13,417 messages from 15 people in OrgA over 4 years and 15,928 messages from 9 members of OrgB over 2 years. The top 30 experts in a range of topics were manually identified and the HITS (Hypertext Inducedt Topic Selection) algorithm performed 52% precision for OrgA with 38% recall. For OrgB precision was 67% with 33% recall.

In contrast, Autonomy[2] analyzes users' search and publication histories to determine concepts that are indicative of their expertise. Yenta (Foner 1997) and Tacit KnowledgeMail[3] determine user expertise from email message traffic. MIT's ExpertFinder (Vivacqua 1999) instruments software library usage to determine expertise level. Referral Web from AT&T (Kautz et al. 1997) provides access to experts across an enterprise or community, aiming to make the basis for referral transparent to the user. It generates social networks based on bibliographic information and supporting context to deduce actual experts and associated referral paths. U.S. West's Expert-Expert Locator (Streeter & Lochbaum 1988) also finds experts across an enterprise.

Abuzz's Beehive[4] is one of many systems that provide an on-line community environment to support question/answer dialogues between users and registered "experts." Users can learn from other user's question/answer dialogues posted under specific topics 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 ontology, which could be browsed by users to find questions/answers similar to their own question. If users did not find a related question they were referred to an expert. The emerging on-line commercial

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

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

[3] Tacit Knowledge Systems' KnowledgeMail (*http://www.tacit.com*)

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

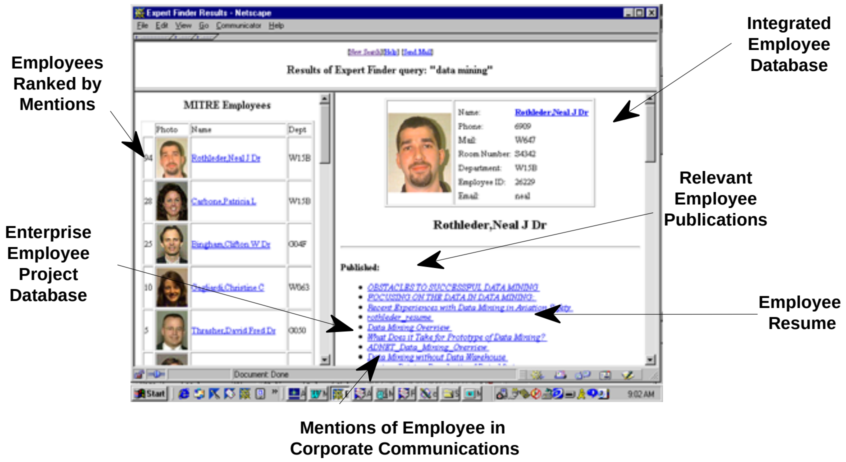
systems attempt to also track each experts' performance; and the general trend is to use user ratings and experts response times as a basis for measuring competence.

Essentially, social filtering is used to qualify the level of expertise of registered experts. Recommender systems have been used successfully to recommend movies, music, books, videos, web pages and usenet news. As such systems suffer from the cold-start problem (no initial ratings), exacerbated also by the fact where there is a mismatch between the number of experts and users. In some cases experts outnumber users, discouraging experts' participation or affecting revenue. In other cases, there is a dearth of experts (or qualified experts) and users become frustrated because of poor response times or low quality answers. While these systems (e.g., XperSite.Com[5]) present interesting expertise management paradigms a number of core problems remain including representing and measuring an expert's qualifications, as well as matching questions to the appropriate experts.

4.1. MITRE's Expert and Expert Network Finder

Mattox, Smith, and Seligman (1998) describe MITRE's Expert Finder which helps to fill this gap by mining information and activities on MITRE's corporate intranet related to experts and providing this in an intuitive fashion to end users. Figure 1 illustrates the initially created prototype in action. In this example, a user is trying to find data mining experts in The MITRE Corporation. When the user searches using the term "data mining," the system ranks employees by the number of mentions of a term or phrase and its statistical association with the employee name either in corporate communications (e.g., newsletters) or based on what they have published in their resume or document folder (a shared, indexed information space). Integrated with MITRE's corporate employee database, employees are ranked by frequency of mentions, pointing to sources in which they appear. Expert finder achieved its original objective, which was to place a user within one phone call of an expert. In empirical evaluations, in spite of the fact that human agreement regarding expertise is surprisingly low (60% or less), over 40% of the experts returned by Expert Finder are judged by humans as experts (a measure of "precision"). Expert Finder also finds about 30% of all the experts identified by human experts (a measure of "recall") (Mattox, Maybury and Morey, 1999).

[5] XperSite.com (*http://www.xpersite.com/*)



Results of Expert Finder query: "data mining"

MITRE Employees

	Photo	Name	Dept
94		Rothleder,Neal J Dr	W15B
28		Carbone,Patricia L	W15B
25		Bingham,Clifton W Dr	G04F
10		Gagliardi,Christine C	W063
5		Thrasher,David Fred Dr	G050

Name:	Rothleder,Neal J Dr
Phone:	6909
Mail:	W647
Room Number:	S4342
Department:	W15B
Employee ID:	26229
Email:	neal

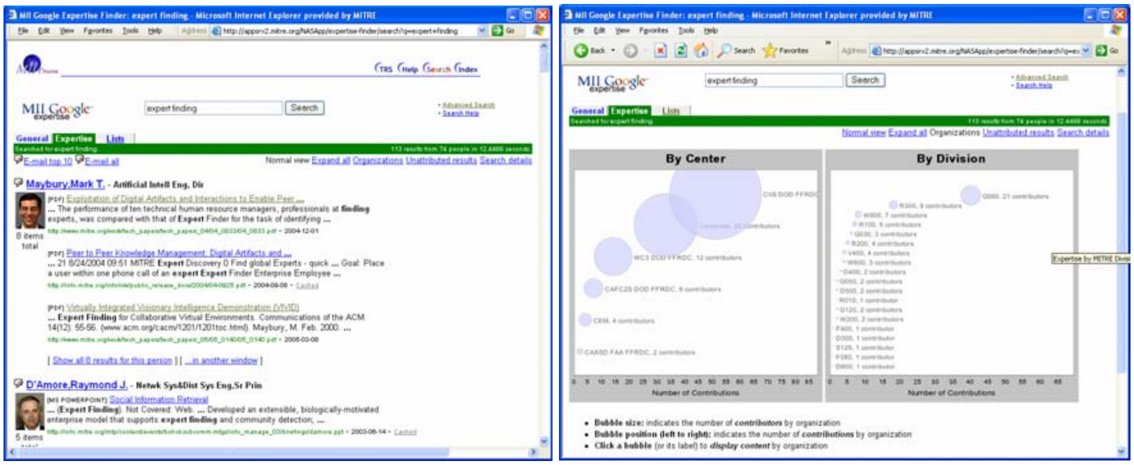
Rothleder,Neal J Dr

Published:

- OBSTACLES TO SUCCESSFUL DATA MINING
- FOCUSING ON THE DATA IN DATA MINING:
- Recent Experiences with Data Mining in Aviation Safety
- rothleder_resume
- Data Mining Overview
- What Does it Take for Prototype of Data Mining?
- ADNET_Data_Mining_Overview
- Data Mining without Data Warehouse

Figure 1 MITREs Prototype Expert Finder “Data Mining” Example

Figure 2 illustrates the operational prototype that MITRE deployed corporately based upon experience with the initial prototype. Engineers leveraged a corporate deployment of Google to index content. Users search using a simple keyword interface shown in the left of Figure 2. In the example, a user searches for “expert finding” and is returned the top ranked experts in accordance with evidence from public documents, communications (e.g., listserv contributions), project time charges and so on, which are shown below each expert. This both allows for validation of expertise as well as access to the expert’s artifacts. A user can select an expert or use an “email top 10” or “email all” link to send a note to the experts. Note next to the “expertise” tab is a “lists” tab which allows a user to find expert community of interests, for example, from hundreds of listservs. The user can also select the “organizations” link to automatically generate the right hand screen shot in Figure 2 which displays the number of contributions each MITRE division or center (a group of divisions) so the user can visualize expertise distribution across the corporation as measured by volume of relevant artifacts created by individuals and organizations, in this case on the topic of “expert finding”. While not heavily advertised, the expert finder is accessed about 4,000 times each month.


MII Google Expertise Finder: expert finding - Microsoft Internet Explorer provided by MITRE
File Edit View Favorites Tools Help
Address http://apporv2.mitre.org/NASApp/expertise-finder/search?q=expert+finding
Go
TRS Help Search Index
MII Google expertise
Search
Advanced Search
Search Help
General Expertise Lists
Back Search Favorites

expert finding

Searched for expert finding — 113 results from 74 people in 12.4466 seconds

E-mail top 10 E-mail all — Normal view Expand all Organizations Unattributed results Search details

Maybury,Mark T. - Artificial Intell Eng, Dir

8 items total

[PDF] Exploitation of Digital Artifacts and Interactions to Enable Peer ...
... The performance of ten technical human resource managers, professionals at finding experts, was compared with that of Expert Finder for the task of identifying ...
http://www.mitre.org/work/tech_papers/tech_papers_04/04_0833/04_0833.pdf - 2004-12-01

[PDF] Peer to Peer Knowledge Management: Digital Artifacts and ...
... 21 8/24/2004 09:51 MITRE Expert Discovery 0 Find global Experts - quick ... Goal: Place a user within one phone call of an expert Expert Finder Enterprise Employee ...
http://info.mitre.org/info/public_release_docs/2004/04-0925.pdf - 2004-09-09 - Cached

[PDF] Virtually Integrated Visionary Intelligence Demonstration (VIVID)
... Expert Finding for Collaborative Virtual Environments. Communications of the ACM 14(12): 55-56. (www.acm.org/cacm/1201/1201toc.html). Maybury, M. Feb. 2000. ...
http://www.mitre.org/work/tech_papers/tech_papers_05/05_0140/05_0140.pdf - 2005-03-08

[Show all 8 results for this person] [...in another window]

D'Amore,Raymond J. - Netwk Sys&Dist Sys Eng,Sr Prin

5 items total

[MS POWERPOINT] Social Information Retrieval
... (Expert Finding). Not Covered: Web. ... Developed an extensible, biologically-motivated enterprise model that supports expert finding and community detection; ...
- 2003-06-14 - Cached

Normal view Expand all Organizations Unattributed results Search details

- **Bubble size:** indicates the number of **contributors** by organization
- **Bubble position (left to right):** indicates the number of **contributions** by organization
- **Click a bubble** (or its label) to **display content** by organization

Figure 2 MITRE’s Operational Expert Finder

4.2. Locating Expert Networks

In other research MITRE investigated the discovery of communities of practicing experts via a prototype called XperNet. MITRE Technology Centers conduct applied research in a number of technology areas related to the sponsor's mission. As such they often partner directly with project departments and form teams with diverse but complementary skills and problem knowledge. Organizationally, staff working related technologies and problems can be modeled as social networks that form the basis for abstracting expertise. According to Ackerman et al. (1999), expertise networks can be defined as

> *… specializations of an organization's social network. They consider not only how people are socially arranged but what expertise they have and trade.*

XperNet is designed to extract expertise networks. XperNet uses statistical clustering techniques and social network analysis to glean networks or affinity groups consisting of people having related skills and interests. Networks are extracted from various work contexts or activities such as projects, publications, and technical exchanges. In the first generation XperNet architecture, Figure 3, expertise signatures are extracted from a variety of sources including published documents, public share folders, and project information. Higher levels of expertise are associated with features such as document authorship, explicit reference or citation, network centrality, personal web pages, and project membership. Lower expertise levels reflect fewer expertise indicators and possibly counter-indications such as being a member of the administrative staff.

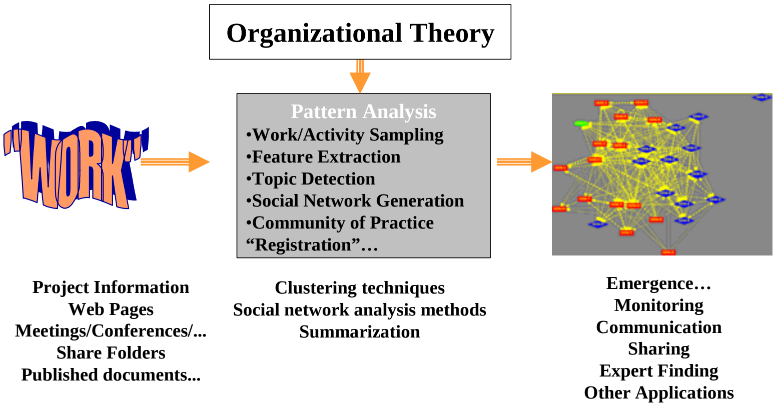


Figure 3 XperNet Computational Architecture

Expertise signatures are input to a simulated annealing clustering algorithm that uses a dynamic merging scheme to collapse lower-level clusters into larger groupings-- the core cluster. The network expansion algorithm "grows" the network by using project and other organizational information to augment expertise ratings and identify additional cluster members. In some cases, members that have complementary skills or serve in supporting roles are incorporated into the cluster. Expertise indicators can be used to indicate expertise ratings for each network member.

The current XperNet algorithm is under evaluation at MITRE's Information Technology Center (http://itc.mitre.org). Initially, expertise networks, generated from user surveys, were compared to automatically generated expertise networks. We found that four nominated core expertise areas (collaboration, knowledge management, advanced instructional training, and language processing) had a strong correlation to existing Center departments, however, in addition many experts from organizations outside of the Center were included, making explicit the informal organization. In empirical precision/recall evaluations comparing human to XperNet performance (Maybury, D'Amore and House, 2000a), approximately 70% of the top ten automatically identified experts were in the manually identified list (precision). In addition, over 40% of the experts included in manually described expert communities were included by XperNet in its list of experts (recall).

Planned research is focused on developing a framework for modeling communities to include aspects of membership, role, and influence. Related to this, a sort of "social semantics" (e.g., expertise, activism) is essential to characterizing community behavior within a specific domain. To motivate this, we have examined problems like Y2K remediation and are currently looking at a specific instance of public opinion on the use of pesticides for crop eradication. General research is focused on the underlying content analysis and social network analysis methods.

5. Performance

Assessing the performance of expert finding tools should take a multidimensional tact. Of course it is important that the system actually be able to find experts. Accordingly technical performance measures such as those described in the previous section (e.g., the precision and recall of a returned expert list) are important. One key to comparing and contrasting systems is a common data set – lists of experts and sources from which that expertise can be inferred. Unfortunately very few organizations assess the performance of their expert finders much less benchmark them against a standard data set or expert finding task. Fortunately, the Text Retrieval and Evaluation Conference (TREC) Enterprise track evaluated both email search and expert search (Craswell et al. 2005). In the latter task, 9 groups participated in the first expertise search task which sought to find experts from 331,037 documents retrieved from the World Wide Web Consortia (W3C) (*.w3.org) site in June 2004. Given 50 topical queries, find the list of W3C people who are experts in that topic area given a list of 1092 candidate experts. Ten training queries were provided. The Mean Average Precision (MAP) of the best system was .275 MAP. MAP is the mean of the average precisions over a set of queries after each document is retrieved. This measure gives better scores to techniques that return more relevant documents earlier. US, European and Chinese organizations participated. Results from TREC are displayed in Figure 4. Unfortunately the commercial solutions described in the next section have not yet been assessed against this benchmark.

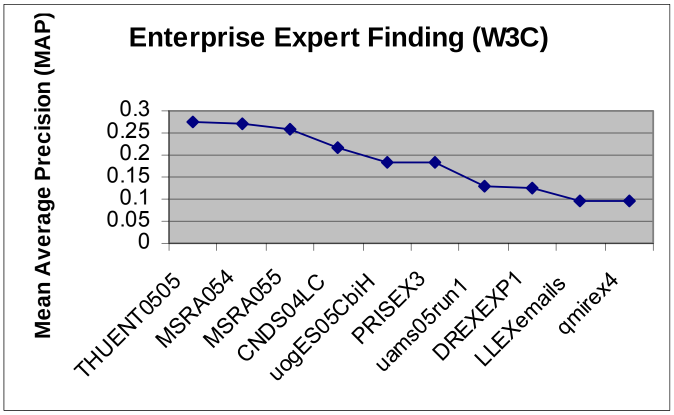


Figure 4 TREC Expert Search Task

In addition to technical measures, however, other measures of assessment can be more important for an organization. For example, benefits beyond speed and quality of retrieval might include:

- *Expert Disclosure*: Does the appearance of expert finding services encourage experts to publish expert profiles or their expert content?
- *Time:* How quickly can individuals find experts or expert knowledge sources?

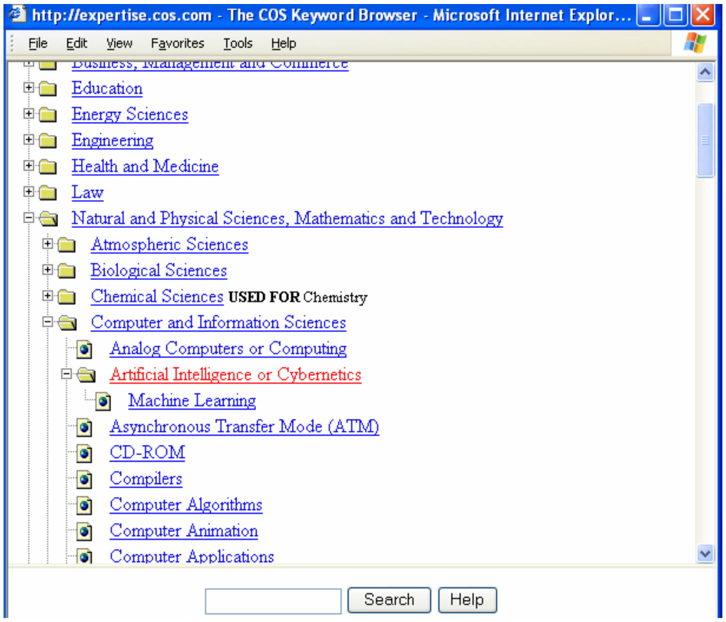
- *Knowledge Searching:* Does the availability of an expert finder increase the amount of knowledge discovery events by end users because they believe they can find answers to their knowledge needs?
- *Knowledge Stewardship:* Does the designation of experts or their increased visibility to staff encourage knowledge sharing?
- *Enterprise Awareness:* The insight the enterprise gains into its staff competencies in terms of areas of expertise, size and depth of staff in those areas.

6. Commercial Services

We distinguish between on-line services that connect up individuals with experts and software systems that an enterprise can purchase or lease to better leverage their own experts. We consider services in this section and software systems in the next section.

A multitude of expert finding services have found their way onto the web. While in some cases technologically similar to match making services for commerce (e.g., E-Bay), relationships (e.g., E-harmony) or social networking (e.g., Linked-In), here we focus on services that are aimed at discovery of professional services or expertise.

One of the largest on-line expert location services is Community of Science (www.cos.com). This site contains profiles of more than 480,000 experts from over 1,600 institutions worldwide. A user can search for experts by keyword, location, name, or institution. They can also browse a hierarchically organized expertise taxonomy as illustrated in Figure 5 to find the kind of expert they need. Analogously, Profnet (www1.profnet.com) was created in 1992 and is used by thousands of journalists from 4,000 news organizations to find experts in a broad variety of domains. It currently sports 600-700 queries per week for 10,000 users. Finally, Expert Witnesses (www.expertwitness.com) is a site used by the legal community to find experts who can consult or testify on a broad range of subjects.


http://expertise.cos.com - The COS Keyword Browser - Microsoft Internet Explor...
File
Edit
View
Favorites
Tools
Help
Search
Help

Business, Management and Commerce
Education
Energy Sciences
Engineering
Health and Medicine
Law
Natural and Physical Sciences, Mathematics and Technology
Atmospheric Sciences
Biological Sciences
Chemical Sciences **USED FOR** Chemistry
Computer and Information Sciences
Analog Computers or Computing
Artificial Intelligence or Cybernetics
Machine Learning
Asynchronous Transfer Mode (ATM)
CD-ROM
Compilers
Computer Algorithms
Computer Animation
Computer Applications

Figure 5 Community of Science (www.cos.com)
Taxonomy Browse and Keyword Search

In addition to services for finding experts, services for answering questions have also proliferated. For example, Abuzz Beehive (described above) was purchased by The New York Times Company in 1999 in order to create innovative online communities around specific content areas on the Web, to provide answers to questions on topics such as wine and news on sites such as www.winetoday.com, Boston.com, and www.nytoday.com. Vistors to these sites register and enter profiles which allows them to post questions (e.g., on wine, Boston, New York) to the Beehive. The Beehive routes these questions to the people in the network who can best answer them based on their profiles. The more one interacts with the Beehive, the more accurate it becomes in routing individual requests.

In contrast to this commercial service, Wondir (www.wondir.com) is a free, open site for Frequently Ask Questions (FAQs). Wondir hosts over 200,000 experts, has 2 million questions on line, and processes about 8,000 questions and 11,000 answers per day. Similar sites include All Experts (www.allexperts.com) and Google (answers.google.com/answers).

Finally, technology is emerging which attempts to automatically answer questions without a human in the loop. These systems perform question answering, document retrieval, information extraction, and answer ranking to provide automated, real-time answers to users from either public collections (such as the web) or private ones. They are sometimes applied to enhance web sites of organizations. Examples include Language Computer (www.languagecomputer.com) and MIT's START (start.csail.mit.edu).

7. Commercial Tools

Commercially available solutions to expert or expertise finding have appeared (and disappeared) in the marketplace. While many on line expert finding services have emerged (e.g., for expert witnesses in trials, for domain expert for journalist to interview), this report focuses on commercial software that an enterprise can deploy to support finding their own or other organizations' experts.

Table 1 provides an overview summary of key expert finding tools in the market place. Currently available commercial tools are listed in the rows (e.g., TACIT, AskMe) and key product features are indicated in the columns grouped into key feature clusters including:

- **Sources** processed to determine expertise. Source can include be self declarations of expertise and/or artifacts created by an expert such as their electronic mail, documents (e.g., Microsoft Word), briefings (e.g., Powerpoint), resumes, web pages or databases (e.g., Peoplesoft, Oracle Financials, project information). Whereas expertise may be stated manually by a user in an expertise profile, it may also be automatically extracted from communications or artifacts of the expert, reducing burden on the expert and ensuring up to date expert profiling. Sources can also be more behavioral, such as the content of the searches an expert makes on the web or against a specialized collection. Effective privacy management can enable expertise profiling to encompass both published and unpublished sources. This feature is captured in Table 1 under system properties.
- **Processing**: The kind of processing the software performs such as automatically ranking the level of expertise of a particular individual on a particular topic, extracting entities from a text such as the people, organizations, locations, materials, or topics that are present in documents authored by the expert. Also systems might perform some level of social network analysis (e.g., looking at email communications, co-authorship of documents, co-work on projects) to determine relationships among experts or communities of interest. Some systems process only English text and in others they handle foreign languages. The number in the table indicates how many languages the product supports. Finally, author identification means going beyond observing when an author publishes a document and beyond extraction of metadata from a document's author properties to performing language processing on document content to automatically determine authorship. This is an important but largely unaddressed capability.

Table 1 Expert Finding Systems

Expert Finding Tools

Capability

■ Full
▨ Partial
□ None

PRODUCT	Sources								Processing					Search				Results			System		
	Self declaration	Email	Documents	Briefings	Resumes	Web pages	Databases	Behavior/searches	Ranking	Entity Extraction	Social Net Analysis	Foreign Language (#)	Author Identification	Keyword	Boolean	Natural Language	Taxonomy (Browse)	List of Experts	Related Documents	Related Concepts	Interoperability	In Operational Use	Privacy
TACIT	■	■	■	■	■	■	■	□	■	▨	□	▨ 2	□	■	■	□	□	■	▨	▨	■	■	■
AskMe	■	■	■	■	■	■	■	□	■	□	□	▨	□	■	■	■	■	■	■	□	■	■	■
Autonomy	■	■	■	■	■	■	■	▨	■	■	□	■ 70	□	■	■	■	■	■	□	□	■	■	▨
Endeca	■	■	■	■	■	■	■	▨	■	■	□	■ 250	□	■	■	■	■	■	□	□	■	■	□
Recommind	▨	■	■	■	■	■	■	▨	■	▨	□	■ 200	□	■	■	▨	▨	■	■	▨	■	■	■
Trivium	■	▨	■	■	■	■	■	□	■	■	□	▨ 6+	□	■	■	▨	■	■	■	▨	■	■	▨
Entopia	□	■	■	□	□	□	□	■	■	▨	■	▨ 6	□	■	□	■	▨	■	■	■	■	■	□

- **Search**: The kind of retrieval supported by the tool, including keyword, Boolean (e.g., AND, OR, NOT) or proximity, natural language query, and/or taxonomic browsing of the content.
- **Results**: The way in which results are presented to the user, such as in ordered lists of experts, lists of documents or artifacts produced or used by experts, and concepts related to the topic of expertise the user is searching for.
- **System Properties**: The kinds of interoperability support provided by the tool (e.g., interfaces to staff directories, document management systems, human resource and financial systems), if it is broadly in operational use, and the kinds of privacy support provided by the tool. European and Asian privacy laws require support for up front consent and opt-in for users whereas under US privacy laws organizations own their own data and are less constrained. Nevertheless, processing of personal communications such as email or instant messaging (the latter of which no current tool supports) are often of concern to employees, so this is identified as an important system property.

Each tool is assessed on these features as providing no capability, partial support or full support. Detailed product descriptions then provide richer description of each tool and the features it provides. In addition, each product write up includes a description of system deployments (the type, size, and complexity of installed systems) and cost (license, maintenance, or renting).

7.1. TACIT ActiveNet™

Head Quarters:	Palo Alto, CA
Founded:	1997 by David Gilmour, formerly of Giga and Lotus
Employees:	under 30
Revenues:	Privately held
Contact:	Rick Hartung, Regional Sales Manager, (201) 848-5455, Rick.Hartung@tacit.com.
Home Page:	http://www.tacit.com

TACIT ActiveNet™ is a unique, patented software solution that addresses the problem of organizational disconnectedness. By automatically understanding enterprise activity in real-time, ActiveNet enables employees throughout the organization to connect with one another on key topics and speeds the organization's ability to solve problems and address issues.

Sources: While TACIT does enable experts to self declare expertise, its primary strength is the automated processing of expert's email and artifacts they produce such as documents or briefings as well as descriptive material about themselves such as resumes or web pages. In TACIT's internal processing model, a source is anything with a date/time, author, and content. Because of privacy concerns (the desire to preserve anonymity and not be a surveillance system), TACIT does not perform "behavioral" processing, such as analysis of search behavior, project charging activity, or document or software repository access. TACIT's privacy management services enable TACIT to take advantage of both published and unpublished expertise (e.g., in private communications such as email).

Processing: TACIT automatically ranks the level of expertise of a particular individual on a particular topic, extracting frequencies of nouns and noun phrases and context of surrounding words (e.g., "nicotine patch therapy") from unstructured text and the date/time of their appearance. Thus while not classifying extracted phrases into semantic entities such as people, organizations, or locations, TACIT extracts and synthesizes linguistic units into "topics", which are proxies for semantic distinctions. For example, TACIT automatically detects the difference between "clinical trials" and "criminal trials" by observing that these two phrases cluster into different uses. TACIT does this based on actual utilization of language within all or part of an organization, allowing, in effect, the automatic generation of dynamic taxonomies that are derived at any level of aggregation within an organization. This differs from static taxonomies, which require explicit human editing and manipulation which are neither scaleable nor timely/current.

TACIT automatically creates a profile for users (from analysis of (non-confidential) email and documents) and allows them to modify that so they have both a public profile, which is accessible to others, as well as a private profile that will match queries of expert seekers but not reveal their identify to seekers unless they release their identity following an alert of the interest. Users can attach "related materials" to their profile, such as a resume or list of expertise areas which would then match queries. TACIT does not perform social network analysis (e.g., co-authorship of documents, co-work on projects) to determine relationships among experts or communities of interest. However, since TACIT processes email and content it could determine from content that someone knows something about a company and from a domain name of an email (e.g., microsoft.com) that that person has had communication with that company. TACIT does not perform authorship identification (although of course it can use metadata about an individual). In addition to English, the system can process French and German sources.

Search: TACIT supports keyword and Boolean (e.g., AND, OR, NOT) query. While users can search for phrases, it does not perform full natural language processing of the query and/or support taxonomic browsing of the content.

Results: Results are presented in ordered lists of experts ranked by a confidence rating based on frequency and recency of content. Because of TACIT's sophisticated privacy model if the search is on "public" profiles, results are presented as a 0-100% match on a person. If the search matches a private profile, it will return a list of relevant topics not people and the individual(s) with the matching private profile will be alerted of the interest and can elect to reveal their identity or not. The system offers to send an email to the top N experts with private profiles, protecting privacy of recipient of email. There is no blind copy so that only the recipient knows, that is there is a "tacit" email. In this manner TACIT can take advantage of both published and unpublished expertise.

System properties: TACIT provides a connector API, and supports MAPI, LDAP, Microsoft Exchange, Lotus Domino, Microsoft SQL and Oracle 9i, J2EE Compliant Application Servers, BEA WebLogic and IBM's websphere (IBM uses TACIT on their web site). They provide interoperability with search engines on a one off basis and built a Groove connector for the federal government. TACIT's "connector toolkit" enables developers to integrate essentially any content source into the product.

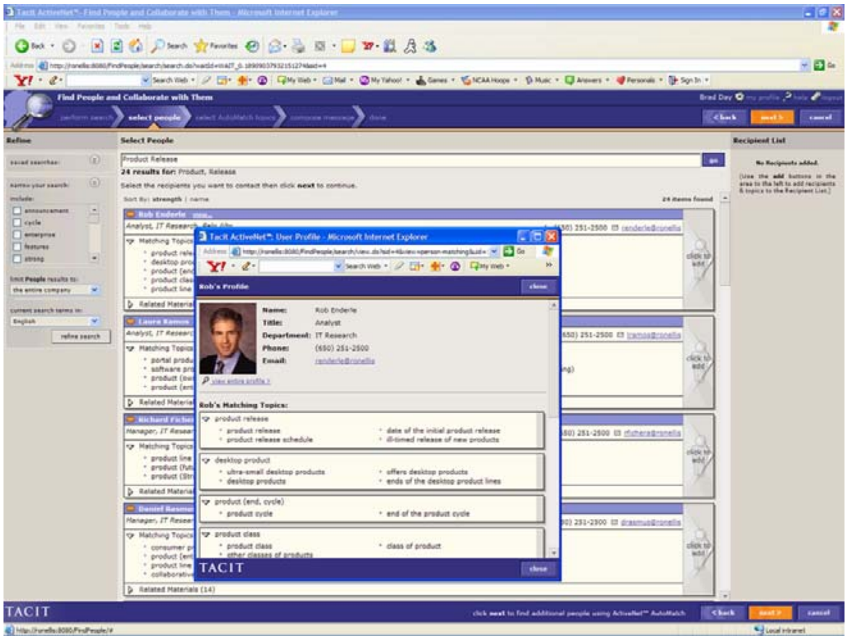
TACIT provides application skins on top of their core engine for specific functions such as pre-procurement, six sigma, and R&D applications of expert finding. Profiles are built within the first week, and full scale deployment typically takes a month or less (2-3 months in complicated situations).

While the system has not been benchmarked against any standard collections such as the NIST collection, it has been applied to both French and German.

Deployments: Deployments include Fortune 250 companies such as GlaxoSmithKline, Lockheed Martin, Northrop Grumman, Morgan Stanley, and the U.S. Government. Deployments range from several thousand to one hundred thousand seats. In the future they plan to explore reverse auction search or expert search such as desktop to desktop (see www.illumio.com). Figure 6 illustrated a display of a search for experts on "product release" in TACIT ActiveNet™.

Cost: TACIT's pricing has 3 components: licenses (searchers and profiles), services for deployment and annual support. Licenses for 5000 searchers/profiles cost approximately $500,000, for 10,000 searchers/profiles: approx. $1,000,000, and for 20,000 searchers/profiles: approx $1.4m, each inclusive of services and 1st year of support. The profiles can be bought separately and range from 0-5000 ($100,000) to 5000-10,000 ($200,000), 10,000-20,000 ($300,000) and 20,000+ ($400,000). The searchers (seats) will range from approx. $260/seat for under 2500 to under $25/seat at over 40,000 seats. Thus, the total "blended" (all included) costs for 2500 seats/profiles to over 40,000 seats/profiles will range from approximately $200 for 2500 seats/profiles, to under $50/seat for over 40,000 seats/profiles. The annual support is 18% of the license cost. Services are $2000/day.

Discussions/Differentiation: TACIT's patented automated profile creation, direct addressing of email, privacy model and support for anonymity differentiate it in the marketplace. TACIT is in discussions with Microsoft, Yahoo, and Google as they seek to broker connections among people. While TACIT notes that Microsoft has announced plans to provide services like this, they feel strong about their 13 granted and 11 pending patents for ActiveNet™.



Product Release

24 results for: Product, Release

Select the recipients you want to contact then click **next** to continue.

No Recipients added.

(Use the **add** buttons in the area to the left to add recipients & topics to the Recipient List.)

Name: Rob Enderle
Title: Analyst
Department: IT Research
Phone: (650) 251-2500
Email: renderle@ronello

Rob's Matching Topics:

- product release
  - product release
  - product release schedule
  - date of the initial product release
  - ill-timed release of new products
- desktop product
  - ultra-small desktop products
  - desktop products
  - offers desktop products
  - ends of the desktop product lines
- product (end, cycle)
  - product cycle
  - end of the product cycle
- product class
  - product class
  - other classes of products
  - class of product

Related Materials (14)

click **next** to find additional people using ActiveNet™ AutoMatch

Figure 6 TACIT Active Net™

7.2. AskMe

Head Quarters:	Seattle, WA (Branches in Schaumburg, IL and India)
Founded:	1999
Employees:	65
Revenues:	privately held
Contact:	Dan Teeter, Director of Sales and Solutions, DTeeter@AskMe.com, (425) 564-9017
Home Page:	www.askmecorp.com

Sources: AskMe automatically processes a range of sources including documents, email, and external publications. Sources are sent to FAST (Fast Search and Transfer, www.fastsearch.com) to select keywords (proprietary, but handles many formats). Users can also manually add in expertise keywords in their profile.

Processing: AskMe distinguishes between Auto profiling and Dynamic Profiling. *Auto profiling* is expertise identification based on the mining of keywords either directly through AskMe or a third party repository (e.g. email or document repository). In contrast, *Dynamic Profiling* is the identification of expertise based upon the solutions you provide and documents you author either directly into AskMe or in a third party repository. For example, if you publish a document(s) on a particular subject, when other users perform queries that match those documents, AskMe will identify not only the content of the document as a source of information but your profile will appear as a potential knowledge provider on the subject as well. Dynamic Profiling is supported by indexing document content using FAST which supports 250 file types (special document formats like CAD are not handled). Expertise validation is based on the pedigree of information sources and expert qualifications/certifications in metadata. AskMe also supports routing information or approval loops to validate information. It has a rich Question and Answer feature set and has community pages for Community of Practices (CoPs) and a related feature set. AskMe does not automatically determine document authorship. AskMe does not perform social network analysis or “behavioral” processing, such as analysis of search behavior, project charging activity, or document or software repository access. AskMe supports English and major European languages (French, Spanish, German, etc.) but currently does not support characters for Chinese or Japanese. An autotaxonomy tool maps documents into classes and so can show suggested taxonomy classification.

Search: AskMe’s web based interface allows users to search using natural language, Boolean, or fields based search (e.g. drop down boxes containing different ontology’s or taxonomy classifications) Users can search experts by skills, experience, project history, certifications, an so on.

Results: AskMe returns a list of experts or related documents. An automated metadata tool can be displayed along with the results.

System Properties: AskMe is interoperable with a broad range of systems including LDAP (e.g., expert’s title, department, security information), Email (Outlook, Lotus), portal integration software (Plumtree, Oracle), document management systems like Documentum, Livelink, Hummingbird), search engines (Verity, Autonomy, Google), project management tools (ERoom), human resources and skills databases (e.g., PeopleSoft for training and educational information), or other databases such as patent or publication databases. It is not currently interoperable with Oracle Financials but can be modified according to the user requirements.

With respect to security, users have complete control over their individual profiles. Content can be controlled at the system administrator or community level. Content tagging enables the restriction of certain kinds of content and user access for sensitive information. Finally, business rules can be configured to route sensitive information through an approval loop prior to being published to the knowledge base.

Deployments: AskMe's privacy model is based on content tagging. For example deployments at organizations such as Boeing and Raytheon, Intel, Procter and Gamble, Prat Whitney missiles use metadata to classify data based on ITAR for export control. Business rules can be applied to control who has access to what type of information and how it is disseminated.

Cost: Initial implementation for a pilot ranges from $50k to $150k depending on the size and scope as well as the level of integration and customization required. AskMe highlights the value returned with deep process integration using available modules that can sit on top of the AskMe platform or be deployed as stand-alone applications. Several modules are available that can support complex stage gate process workflows for concept development, best practice vetting and approval, and proposal writing. Customers can also build custom applications on top of the existing platform to meet specific business needs and requirements.

Discussion: Other differentiating features of AskMe are its support of community spaces, workload management and subscription and notification based on profiles. Experts can easily publish FAQs and AskMe claims a methodology for having the right solution. They support rating of content (e.g., assessing $ vs. time saved).

Differentiation: AskMe views their product distinct from search products such as Autonomy or Entopia which can do metadata extraction and autoprofiling of documents. Figure 7 illustrates the AskMe architecture which includes a presentation layer, an API layer, and a range of information sources.

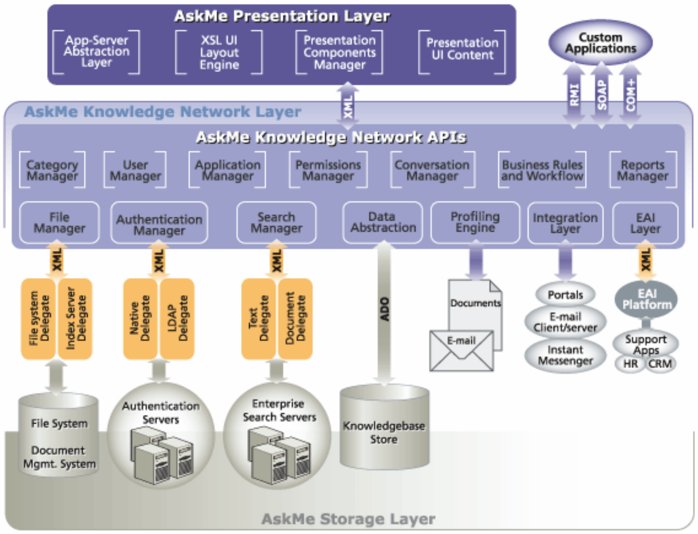


Figure 7 AskMe Architecture

7.3. Autonomy IDOL K2

Head Quarters:	San Francisco, CA and Cambridge Business Park, England.
Founded:	1996
Employees:	many
Revenues:	$1.5 billion
Contact:	Brant West (703) 289 8850. brantw@us.autonomy.com
Home Page:	http://www.autonomy.com/content/Products/K2_Developer/index.en.html

Note: Verity and Autonomy merged in December 2005 and their legacy products K2 and IDOL are merging in the new product, IDOL K2.

Sources: Autonomy automatically processes any text (documents, resumes, web pages) or electronic mail (Microsoft, Lotus). It also analyzes user access to information and applications.

Processing: Autonomy's concept based search utilizes a language independent combination of Bayesian and Shanahan information theory. It includes entity extraction tools (currently Autonomy Eduction which will be combined with IE from the K2 tool). Autonomy supports the relation of many different types of operations beyond keyword search. It tracks over 300 operations that users can perform (e.g., document search, expert portal search, information push, clustering/categorizing). The system could be tuned to particular types of content (e.g., resumes, web pages) for expert finding, but is not. Autonomy does not automatically determine document authorship (uses metadata associating a person with a document) and does not perform social network analysis.

Search: Autonomy supports keyword, Boolean, and concept search and can cluster returns. The tool supports over 70 different languages because their method is mathematically, not keyword, based. It can use an existing taxonomy which it can populate with an index and also automatically create a taxonomy clustering which a user can access via a portlet.

Results: Autonomy returns a list of individual experts based on things they are locating and retrieving. The user can then select people to obtain contact information. Relevance ranking is displayed as a percent based on the volume and relevance of search by a user. It does not currently but could include (manual) forms based input or document analysis.

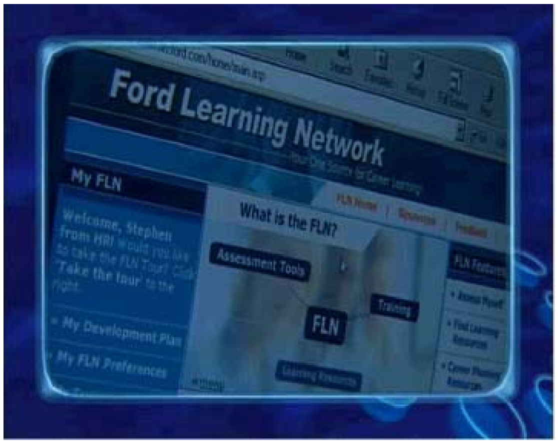
System Properties: To support privacy Autonomy can turn on or off an agent which searcher different repositories. It also can restrict searches in particular communities or on particular topics by individuals. Its MAP security model integrates with ACL.

Autonomy has over 200 out of the box connectors for common applications such as email and attachments, LDAP, web (http), document management (e.g., Documentum), databases (e.g., Peoplesoft, Oracle, ODBC), audio and video. It has a toolkit, Omnifetch, which can be used to build connection to a new source.

Deployments: Approximately 20% of Autonomy's customer base is government, 10-15% of that US federal government. For example, it's used by the Department of Energy to monitor outbound email and apply a rule set to ensure no inappropriate information is sent. Its also been applied at the Joint Special Operations Command (JSOC) at Ft Bragg, Secret Service, other intelligence agencies, Associated Press, CNN, New York Life, Coca Cola, and GlascoSmithKline.

Cost: A typical deployment costs around $300,000. Since a lot of the functionality is automated, cost is often driven by the complexity of the user interface UI (e.g. how many functions, # of repositories want to index against.) On average pretty good size production systems are up and running in a couple of months. Autonomy did a 2 TB implementation last late summer/fall for COCOMs about 2 months ago.

Discussion: Autonomy claims the best of both worlds from the 2005 merger of Verity's K2 and Autonomy's IDOL, illustrated in Figure 8. Where as Verity K2 took the tact of self declaration of expertise, Autonomy IDOL viewed that expertise would change over time. Autonomy notes Verity's K2 capabilities of taxonomy generation, customization, scripting languages, and fine-tuned keyword indexing have been integrated with the recently rebuilt Autonomy IDOL kernel as a multithreaded engine that provides a 30-35% increase in speed with enhanced scaling and security. Autonomy finds search engines FAST, Endeca, and Google as their typical competitors. They state they differentiate themselves by providing manual and real-time proactive search based on how users interact with the system. Figure 8 illustrates Autonomy IDOL K2 in support of Ford's Learning Network which supports 350,000 global employees.



Welcome, Stephen from HR! Would you like to take the FLN Tour? Click 'Take the tour' to the right.

Figure 8 Autonomy IDOL

7.4. Endeca

Head Quarters:	Cambridge, MA
Founded:	1999
Employees:	350
Revenues:	Undisclosed; 142% annual growth rate 2002-2005
Contact:	jmackay@endeca.com, 703 655 1831(Cell) or 703 860 7545 x 207 (Office) (European Office: +44 (0) 207 375 9333)
Home Page:	http://endeca.com See also: http://endeca.com/byProject/directories.html

Sources: Endeca supports access to both structured and unstructured data and can combine data from multiple sources, while retaining and indexing the explicit and implicit relationships, to provide a unified view across siloed systems. Endeca for Directories gives customers the ability to find the right person for staffing, resource management, and networking.

Processing: The Endeca search engine ingests structured and unstructured data. Their relevancy ranking strategy considers where matches are found (e.g., in title or body) word order, frequency, all/some, and proximity (these are 4 of our 12 relevancy modules that can be selected). Their software includes a named entity (NE) tagger (e.g., people, places, things). Endeca also partners with third party NE taggers (e.g., InXight, Aerotext, Netowl, Attensity) depending upon customer needs. They support foreign language identification of 250 languages. They don't require a taxonomy but will integrate one if it exists. For example, in their IBM Global Services Professional Marketplace expert finding tool deployment, they use the company's own taxonomy and augment it with their automated processing to support deeper taxonomies. Endeca can use past search profile information as a source of evidence for expertise. Endeca does not automatically determine document authorship (it uses explicit tagging of metadata to determine author, time of creation, edit time) and does not perform social network analysis. They rely on external sources to validate expertise or quality (e.g., Endeca's ability to incorporate data from multiple systems and retain the relationships enables taking into account a person's past project work, project documents, and other related activities over time to infer expertise, as at IBM).

Search: Endeca's Guided Navigation user experience helps users search and navigate through listings of people by using key attributes, originating in content ranging from structured directories to unstructured resumes. Endeca claims to be distinct from the word list approach used by Convera, Autonomy, Google, or FAST by allowing *Guided Navigation* (e.g., in a medical domain this might present criteria in dynamic menus for location, people, surgery type) which then help users build a complex query behind the scenes. Their patented Guided Navigation both assists in building complex queries and points out questions and ideas that people may not know to ask – by presenting these options in organized, dynamic menus – based on availability of expert qualifications in the data itself. This aids in discovery and helps ensure that people find the ideal expert based on available personnel, including allowing for who is physically available / unavailable based on changing staffing allocations, geographic considerations (including GIS data), and other hard to anticipate criteria. IBM Global Services rolled out this capability to it's 100,000 consultant expert directory to improve staff deployment and wrote about saving $500MM in 2005 in Business Week and the Wall Street Journal (they call the application the "Professional Marketplace").

Results: Endeca has a deep range of Results options to support specific business options. These include:

- *Granular Adjustable Relevance Ranked List:* Endeca returns lists from previous navigation and search. Lists relevance ranking can be finely tuned to reflect business priorities, including such ideas as: prioritize by geographic or time zone proximity, prioritize by years of experience, prioritize by rate, etc. Rules can also take into account user profile information.

- *Guided Navigation Context:* Results are wrapped in a dynamic, data-driven set of menus that are calculated automatically from the intersections in the underlying metadata. This Guided Navigation context presents every possible refinement of the result set implied by the explicit metadata and any auto-extracted metadata. As a feature in the core Endeca engine, Guided Navigation is all handled beneath the application tier for sub-second performance and rapid application delivery. It also integrates with other features in the engine tier, such as security / role-based behavior, etc.
- *Security/Roles*: Endeca intrinsically supports security and role-based behavior at the core engine tier. This means that result presentation – including availability, relevance ranking, spell correction and navigation menus – is intimately and automatically updated for security/roles. So, it is possible to create fully customized views and behaviors by user roles that straddle all system functions. Some companies use these to encourage rules more appropriate to divisions or likely individual interests (e.g., the selection of preferred experts by region, etc.)
- *Result Details***:** Endeca's index incorporates the relationships between underlying data from structured/unstructured repositories. As such, it can be queried to present synthesized views across those systems in the result list. For instance, a person's entry in a results list might include experience attributes, availability attributes, and unstructured documents from their resume and past projects all in a single entry.
- *Result Spotlighting****:*** Endeca's presentation API includes the ability to create high-level rules that spotlight certain results. An example might be: "Highlight the lowest-cost 3 consultants in India that match any criteria the user searches/filters by" which one might use to encourage discovery / exploration of expertise in new / offshore locations. These rules are created / managed via a web-based interface, so business users retain ultimate flexibility / control**.** These rules can expose any type of information, from special documents matching the rule's criteria to special metadata to static redirect pages. The rules can be ranked and tied to different user profiles. Often this spotlighting engine is used for merchandizing or highlighting new or important information in the context of search results**.**

System Properties: Endeca is interoperable with LDAP, PKI (single sign on), JDBC, ODBC, Exchange Mailboxes, Lotus Notes Mailboxes, Mime message stores, XML repositories, Flat files (delimited or one of 370+ formats including Word, Excel, and PDF), Sharepoint, Documentum, Vignette, Peoplesoft, Oracle Financials, Oracle Label Security (Record releasability can be controlled on a per-record and even sub-record basis) and is JSR 168 compliant (Plumbtree, BEA web logic, IBM Webspheres). It provides no direct native access to other vendor's search engines.

As the inventor of Guided Navigation and related information access techniques, Endeca has built and proven out a new framework to ensure adequate scaling. This includes:

- *Query Traffic*: Endeca powers some of the most highly trafficked applications on the web, including the online retail presences for Wal-Mart, Home Depot and Kmart. The Wal-Mart application, for example, routinely handles 600 operations/second. This is supported by straightforward, standards-based parallel scaling behind a standard HTTP load balancer.
- *Dimensional Scope*: Endeca has deployed with over 20,000 attributes and 10's of millions of records for real-time navigation at places like Arrow Electronics. Scalability in attributes (dimensional scope) also applies to the number of possible values in a single dimension where, for some deployments, this ranges into the millions (for instance, the number of possible geo-coordinates). In contrast to alternative approaches to this problem, performance does not degrade relative to the total number of dimensions in the system.
- *Data Volatily:* Information Access must be able to operate in a rapidly changing world, and updates to enterprise data must be reflected quickly in all facets of the system in order for accurate decisions to be made. At places like Overstock.com for example, Endeca accommodates availability and pricing updates throughout the day in order to reflect the constant changes resulting from inventory fluctuations and product liquidations. At IBM Global Services, the availability of experts is constantly changing. Endeca's data-driven MDEX engine automatically incorporates these changes and delivers them to the dynamic UI without code changes.
- *Hierarchy Depth*: It is critical that a system be able to leverage any and all taxonomies relevant to an organization, such as geography hierarchies, product and content taxonomies, and organization or domain-related classifications. Endeca has handled the integrated navigation of search and analytics results in hierarchies with 100,000's of nodes at customers like Weatherford Systems and John Deere without suffering any degradation to performance.
- *Record Scale*: Handling high-record navigation and search in context presents additional challenges, especially when synthesizing navigation that may span more records than can be kept in a single in-memory representation. Endeca's patented Guided Navigation has solved this problem, enabling interactive user navigation in certain financial services applications across 1 billion atomic records and their associated metadata.

Deployments: Endeca is used at 400+ organizations including IBM, American Express, Harvard, MIT, K-mart, Walgreen, and Wal-Mart, CIA, FBI, and DIA. When users can find the right person, organizations see the benefits of staffing projects more efficiently and maximizing the value of human resources and relationships. IBM reported saving $500M in a single year with their Human Supply Chain solution, running on Endeca technology, because they could optimize staffing decisions for their global consulting business (dubbed the "Professional Marketplace", this has been written about in Business

Week, the Wall Street Journal, and IBM's own Annual Report) BCS (Business Consulting Services, now GBS – Global Business Services) separating people by skills, pay grade, past performance metrics, geography, and so on.

The software enables analysts to see human resource gaps (e.g., if an organization needs a Farsi speaking deployable medic). It has become a standard at DIA (it's an 18-24 month process) and via InQTel, additionally, CIA's Langley has an Endeca site license. It is used by the Intelligence Community metadata working group.

In another deployment, the FBI Foreign Terrorism Tracking Task Force (FTTTF) uses Endeca to access individual I94 data for 170M people today, growing to support the accessing of 300M foreign visitors. The complex program has many agencies involved, CBP collects the data with the help of the TSA and GSA scans the hand-completed forms in an operational center located in Kansas City. FBI analysts query the system, and when the investigation progress, thru the process Choicepoint records can be accessed by Endeca to match residences, property purchases and other related information that may be helpful to law enforcement. For American Express the 3/4 billion documents were joined with their Peoplesoft HR data to help find documents, e.g., enabling users to distinguish documents about Cisco coming from the Mergers and Acquisitions department as opposed to the IT department.

Cost: Most Endeca customer's initial investment is between $250k and $1M . Endeca products are available on the GSA Schedule - a 4 CPU platform cost $168K/CPU. Annual maintenance is typically 17% of the licensing fee.

Discussion: Endeca claims to be the only single source provider that bridges unstructured and structured data. A lightweight, robust, and easy to deploy ETL tool supports the transformation of data doing joins across datasets. Figure 9 illustrates the category organization of results returned for a search on "Zinfandel" as well as "guided navigation", which allows search from metadata from the results set such as price, location, year, and so on.

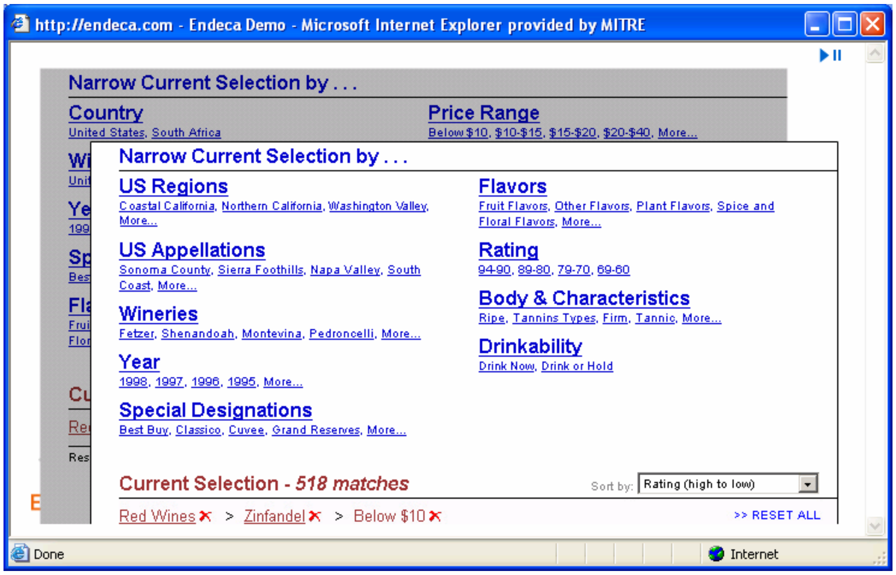

http://endeca.com - Endeca Demo - Microsoft Internet Explorer provided by MITRE
Narrow Current Selection by . . .
Country
United States, South Africa
Price Range
Below $10, $10-$15, $15-$20, $20-$40, More...
Narrow Current Selection by . . .
US Regions
Coastal California, Northern California, Washington Valley, More...
Flavors
Fruit Flavors, Other Flavors, Plant Flavors, Spice and Floral Flavors, More...
US Appellations
Sonoma County, Sierra Foothills, Napa Valley, South Coast, More...
Rating
94-90, 89-80, 79-70, 69-60
Wineries
Fetzer, Shenandoah, Montevina, Pedroncelli, More...
Body & Characteristics
Ripe, Tannins Types, Firm, Tannic, More...
Drinkability
Drink Now, Drink or Hold
Year
1998, 1997, 1996, 1995, More...
Special Designations
Best Buy, Classico, Cuvee, Grand Reserves, More...
Current Selection - 518 matches
Sort by: Rating (high to low)
Red Wines ✕ > Zinfandel ✕ > Below $10 ✕
>> RESET ALL
Done
Internet

Figure 9 Endeca

7.5. Recommind

Head Quarters:	San Francisco, CA (Also facilities in Bonn, Germany, Northern Jersey, and the UK).
Founded:	2000 (by researchers from Univ. of California at Berkeley, MIT, and Brown Univ.)
Employees:	50
Revenues:	Undisclosed, profitable (Privately Held Company)
Contact:	Derek Schueren, derek.schueren@recommind.com, (415) 394-7899, x 223.
Home Page:	http:// www.recommind.com

Sources: Recommind is an enterprise search and categorization company. Recommind's MindServer platform automatically extracts information from back-end enterprise systems - such as document management, records management, customer relationship management (CRM), project and portfolio management (PPM), email, contact, time and billing, intranet, customer support databases, and other systems. It automatically identifies expertise primarily based on the work a person has produced and can also include a self-declared expertise profile. It can process email, say on a public exchange server and has multi-level security capabilities to ensure privacy. Behaviorally it can look at assignments to and time billed to relevant projects, as well as documents authored and edited.

Processing: The MindServer platform is an enterprise-class solution that scales to serve millions of users and search terabytes of data. It is built on the company's patented Probabilistic Latent Semantic Analysis (PLSA) algorithm which performs a statistical analysis of word co-occurrence to create aspect lists. This together with a support vector machine (SVM lite) is used for categorization (which is language independent). Their approach does not rely on static biographies which are likely not current, but rather connects to authored documents and time billed against relevant projects to provide an integrated view of expertise. Some metadata comes from Microsoft Word properties, document management systems or other sources. In some cases metadata can be inferred, for example, Morrison Forester wanted to display client industry in a search, a field that was not in their document management system but could be inferred from a client ID by joining this with the CRM database. Entity extraction from unstructured text is primarily for legal and media entities in English and German language (e.g., people names, company names, presiding judge, jurisdiction, document type, plaintiff, defendant).

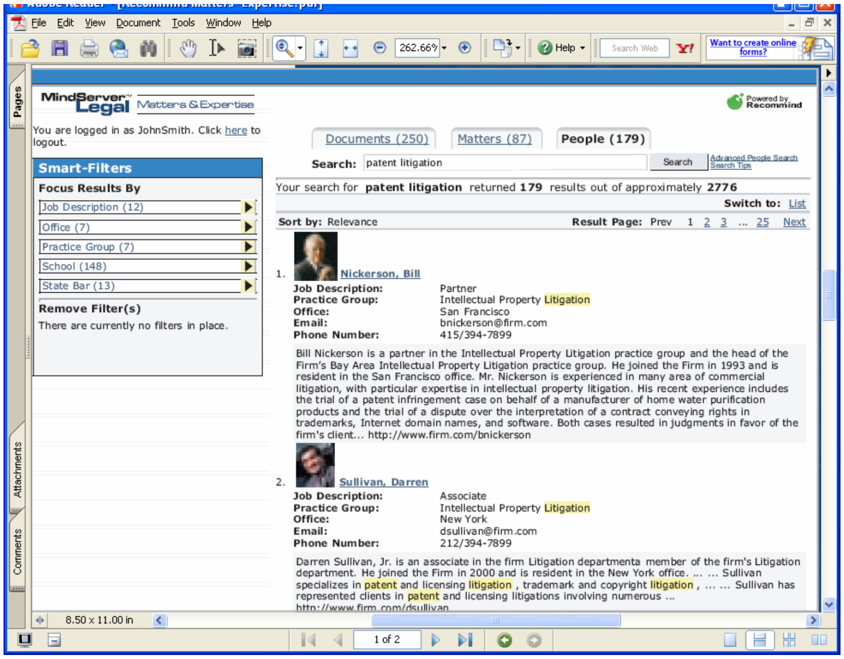
Search: Recommind's technology provides concept-based full text search. In addition to keywords, users can use + or – signs to add or eliminate terms as well as perform proximity searching. While not providing natural language query, it provides "smart filters" which are generated based on metadata (e.g., author, location, practice, industry, matter or project) and can be used as cascades of refinements on the original query. Thus in an example legal deployment, a user searches by keyword and can thus refine the results by selecting a specific industry, then a specific office location, etc. While MindServer does not rely on taxonomies, if a client has a synonym list or taxonomy this can be incorporated (e.g., as was done in medlineplus.com, redlightgreen.org, a library resources portal, and business.gov.au, the Australian government small business portal).

Results: Results are relevancy ranked with metadata based smart filters as described above that allow for further expansion or refinement of the query (e.g., the office, bar-admissions, practice group, or location of a person). You can show relevant documents, people (location, clients, contacts) and matters/projects (client, people assigned, authors, editors, time billed). With concept-based search, MindServer also displays concepts related to the query (e.g., "java" as an island, programming language or coffee).

System Properties: Recommind is interoperable with document management, records management, customer relationship management (CRM), project and portfolio management (PPM), email, contact, time and billing, intranet, customer support databases and other databases. To support privacy, MindServer provides a layer in the software which maps to security in all applications. This allows, for example, the suppression of documents from the results set if the specific user does not have access to it. Many clients also invoke "ethical walls", that is the denial or inclusion of specific people or groups for particular content. The 4.2 version of the MindServer software also allows for additional security rules to be added. The software has an AJAX (Asynchronous Javascript and XML)-based interface.

Deployments: Recommind's customers include Fortune 1000 companies, government agencies, and firms in professional services, financial, pharmaceutical, retail, healthcare, and media firms. Customers include Bertelsmann, DuPont, National Library of Medicine, Cleary Gottlieb, Shearman & Sterling, the Australian Government, and Morrison & Forester. Deployment takes about 8-12 weeks from beta to full implementation. Figure 10 illustrates the Recommind MindServer returning results in a legal corporation for the query "patent litigation".

Cost: Pricing starts at $150K. Per processor or per seat pricing is available. Maintenance is 20% of the license fee.



MindServer™ Legal Matters & Expertise

Powered by Recommind

You are logged in as JohnSmith. Click here to logout.

Documents (250) | Matters (87) | People (179)

Search: patent litigation [Search] Advanced People Search Search Tips

Smart-Filters

Focus Results By

- Job Description (12)
- Office (7)
- Practice Group (7)
- School (148)
- State Bar (13)

Remove Filter(s)

There are currently no filters in place.

Your search for **patent litigation** returned **179** results out of approximately **2776**

Switch to: List

Sort by: Relevance **Result Page:** Prev 1 2 3 ... 25 Next

1. Nickerson, Bill

Job Description:	Partner
Practice Group:	Intellectual Property Litigation
Office:	San Francisco
Email:	bnickerson@firm.com
Phone Number:	415/394-7899

Bill Nickerson is a partner in the Intellectual Property Litigation practice group and the head of the Firm's Bay Area Intellectual Property Litigation practice group. He joined the Firm in 1993 and is resident in the San Francisco office. Mr. Nickerson is experienced in many area of commercial litigation, with particular expertise in intellectual property litigation. His recent experience includes the trial of a patent infringement case on behalf of a manufacturer of home water purification products and the trial of a dispute over the interpretation of a contract conveying rights in trademarks, Internet domain names, and software. Both cases resulted in judgments in favor of the firm's client... http://www.firm.com/bnickerson

2. Sullivan, Darren

Job Description:	Associate
Practice Group:	Intellectual Property Litigation
Office:	New York
Email:	dsullivan@firm.com
Phone Number:	212/394-7899

Darren Sullivan, Jr. is an associate in the firm Litigation departmenta member of the firm's Litigation department. He joined the Firm in 2000 and is resident in the New York office. ... ... Sullivan specializes in patent and licensing litigation , trademark and copyright litigation , ... Sullivan has represented clients in patent and licensing litigations involving numerous ... http://www.firm.com/dsullivan

Figure 10 Recommind MindServer

7.6. Triviumsoft's SEE-K

Head Quarters:	Paris, France (Corp HQ) and Chicago, IL
Founded:	1992
Employees:	30 (25 in France, 5 in US)
Revenues:	$3.5M (grew from $.9M in 2003)
Contact:	Pierre Kergall (312) 674-4560. pkergall@triviumsoft.com
Home Page:	http://www.triviumsoft.com, http://www.trivium.fr

Sources: Triviumsoft's SEE-K (Version 3.5, June 2006) is a skills management tool that automatically extracts skills from any full-text documents including resumes, evaluation forms, project plans or job descriptions in a variety of document formats. They can also draw from corporate Enterprise Resource Planning (ERP) systems (e.g., how much time an employee worked on particular projects). Employee profiles can incorporate a broad range of material such as demographics (automatic extracted from an ERP), training course history, manual entry of self-assessment (skills and history, best or underutilized skills), all of which be used to identify experts. While Trivium has a module for Outlook that processes mail to identify the clusters of words that form islands and topics (based on document and inverse document frequency and word clusters), it has not yet been commercialized. We mark email "partial" in the Table 1 because TIM for Outlook exists and can be easily customized to be integrated into a customer application. The application can be linked to various types of database (e.g., Oracle, Microsoft SQL server).

Processing: Trivium technology is used to map and extract skills from any full-text document to build a preliminary human resource management grid, i.e., skill/competency model, with information linked by content and context. Trivium does not perform social network analysis. Trivium does not perform "behavioral" processing, such as analysis of search behavior, project charging activity, or document or software repository access.

The Trivium Information Mapper (TIM) extracts lists of expressions/words from free texts considering word frequency, co-occurrence across all texts and co-occurrence in each individual text, and "proximity". This enables the detection of both "strong" signals (i.e., high frequency words) as well as "weak" signals (low frequency but very central when it does appear. The TIM technology reads all Western languages (uses a stemmer for English, German, French, Italian, Portuguese and Spanish), and the SEE-K application itself is available in English, French, Spanish and Portuguese. Their "topological" algorithm was created by mathematicians and philosophers. It creates a map - a visual representation of skill distribution based on frequency and similarity. If you have a skill model already and your staff have filled out an on-line self assessment (objects: people and skills) this can be leveraged to populate the application. Trivium technologies represents the link between sets, for example staff and their skills – links are were wealth is. A skills diagnostic or capsule helps identify core/common skills (good for mobility, succession plans, replacement or cross functional team formation) and skill gaps through rare/missing skills identification. Then branches for specific areas of expertise - expert clusters - reveal new areas.

Search: The user can perform full text or Boolean search as well as a structured query based on multiple filter criteria (e.g., search for managers in Chicago with particular certifications or relocation wishes). Queries are saved. The tree structure enables unique search/browse of the skill base. A typical search might start with 2000 experts which could then be narrowed down to 20 via a search, at which point the user could access individual resumes/project reports/or skills to find what you need.

Results: A Capability Tree shown in Figure 11 on the left represents on a single screen an unlimited number of jobs, skills, competencies, employees, training programs, projects and more. The most skilled experts in selected skills appear higher in a rank order which is encoded with color intensity. The screen shot on the right of Figure 11 displays 600 employees and the grouped words in their resumes in the center. In the right display, color is used to display word frequency (the more red a word, the more commonly used in the answers, the bluer, the rarer), the position encodes importance (appearance in the answer) appear toward the center of the map, and words are cluster together because they are co-occurrence.

System Properties: Trivium interoperates with ERP systems (e.g., Peoplesoft). It uses a connector that imports data with a particular frequency to provide an auto-update. It also supports SQL, Microsoft, DB2, Access, Oracle. It addresses privacy at the server level (e.g., their secure hosting of data from the State of Wisconsin), the user level (e.g., employee vs. supervisor), and data layer.

Deployments: Triviumsoft customers are Global 500 organizations across all market segments, including the State of Wisconsin, Airbus, Accenture, Barclays Bank, Suez, CapGemini, Ubisoft, the European Space Agency, Group Credit Lyonnais Bank and France Telecom. They are partners with IBM Global Services, Accenture and Cap Gemini. A Division of the European Space Agency uses their tool to manage rocket launcher engineers. Degremont, a subsidiary of Suez group, uses them for the water treatment plants engineers in which teams of folks across the world need to respond to RFP or bids to fix a water treatment plant.

Cost: Two primary deployment modes are on-site implementation or an Application Service Provider (ASP) hosted in Chicago on secure servers. Trivium also offers rent to own options. For on-site implementations, licenses are based on the number of people managed by the tool with 1,000 employees costing $96,000 list and 10,000 employees $207,000 plus a 16% annual maintenance fee. An ASP is 20-30% less expensive than an on-site implementation, is much faster, and is very flexible (e.g., an organization can do a diagnostic with ASP for a few months and then decide to pursue the product or not). Organizations can rent for 3 months, 6 months, 1 year, 2 years or 3 years. For example, an ASP deployment for 1,000 people on a 12 month rental, the cost is $6/employee/month or about $6k/month. For 10,000 people the price would be $1.32/employee/month or $13.200/month for 12 months. Prices are based on the number of people managed in the application (e.g., a typical deployment might have 1500 people, 300 of whom are managers and 10 HR & IT folks who can do everything).

Discussion: Trivium can be used in both a reactive mode (e.g., what skills, experiences, interests do we have) as well as a tool to assess organizational skills (identifying primary, secondary, and weak skills). Their unique Capability Tree map of skills can be used to do an expert risk analysis, e.g., for retirement, under/over skilled. An organization may want to create a skill model or not. If they do, this data can be used to create skill areas which can drive a survey. An HR diagnostic project, including a skills collection survey of 500-1500 people, takes on average 4-6 weeks. At the end of the survey users can create a Capability Tree and use the application for on-going skills management.

The word Trivium comes from medieval educational theory where the trivium consisted of grammar, rhetoric, and logic. The company was motivated by a former Prime minister of France who recognized that less wealthy in French society often did not have diplomas but they did have important skills, and challenged a top French philosopher to create a system to reveal their skills and knowledge. Michel Serres (member of the French Academy and Professor at Stanford University), set up a core team of experts whose work led to the creation of the Trivium company. Michel Serrres is still a board member of Trivium today.

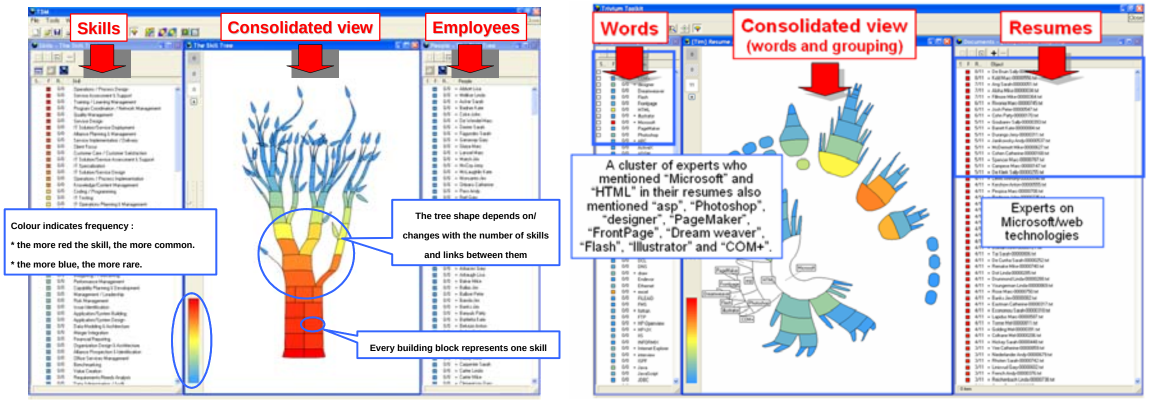

Skills
Consolidated view
Employees
Colour indicates frequency :
* the more red the skill, the more common.
* the more blue, the more rare.
The tree shape depends on/ changes with the number of skills and links between them
Every building block represents one skill
Words
Consolidated view (words and grouping)
Resumes
A cluster of experts who mentioned "Microsoft" and "HTML" in their resumes also mentioned "asp", "Photoshop", "designer", "PageMaker", "FrontPage", "Dream weaver", "Flash", "Illustrator" and "COM+".
Experts on Microsoft/web technologies

Figure 11 Trivium Tree Map Display Employee Skill Frequency and Relationships (left) and a Display of 600 Employees and the Grouped Words in their Resumes (right)

7.7. Entopia Expertise Location

Head Quarters:	Redwood City, CA (Offices in London and France)
Founded:	1999 (Dissolved May 2006)
Employees:	None (See discussion)
Revenues:	
Contact:	Eric Miles, Tel (650) 232-1333, emiles@entopia.com
Home Page:	

Sources: Entopia Expertise Location finds the people in the organization with the most relevant knowledge or expertise to drive collaboration, knowledge-sharing and innovation across the enterprise. The product is based on people's activity as opposed to people's profiles. It is document centric, processing documents, email, and, if available, content from document management systems. Its foundation is a common bus of metadata and keywords (K-Bus). A K-Map or content visualization map can display relations of concepts among documents.

Processing: Entopia creates indexes and metadata related to documents (keyword and semantic) and document library systems (author, modifier, reviewer, frequency of access). Entopia creates an understanding of a topic and related people around it. It focuses on the lifecycle of document(s) and can integrate this with information from Email (e.g., mail sent on different topics). Entopia does not perform "behavioral" processing, such as analysis of search behavior, project charging activity, or document or software repository access. One key differentiator is Entopia's social network analysis which can display a map of interactions by displaying employees as nodes and concepts (e.g., topic or subject) as relations where the size of nodes and lines are based on volume. This can reveal, for example, unique experts as well as which people are gates to communities. The system is Unicode based and supports English, French, German, Swedish (single byte languages) as well as Hebrew and Arabic. When Boeing input AskMe profiles as part of the Entopia algorithm, they were weighted low given the limitations of self declaration.

Search: A user can search using keywords or concepts (e.g., "cell phone sales"). They can display results as keywords or concepts. A box of relevant concepts (in this case “wireless”, “Intel”, “network”, “mobile”) is displayed and can be selected to allow a more narrow search. The user can go to a K-map to see relevant concepts and navigate a tree of related concepts, e.g., “cell phone sales”. Underneath this top concept are subconcepts such as “headsets”, “Nokia”, etc. Thus the user has a kind of hierarchical or taxonomic search. The searcher can also select the top people from expert finder or a source such as Documentum (or a document) which will further restrict the results list.

Results: Can be displayed as a list of relevant experts, related documents, related concepts or a social network. Sources of evidence are weighted (e.g., authors higher than readers or reviewers or email senders) to help rank experts.

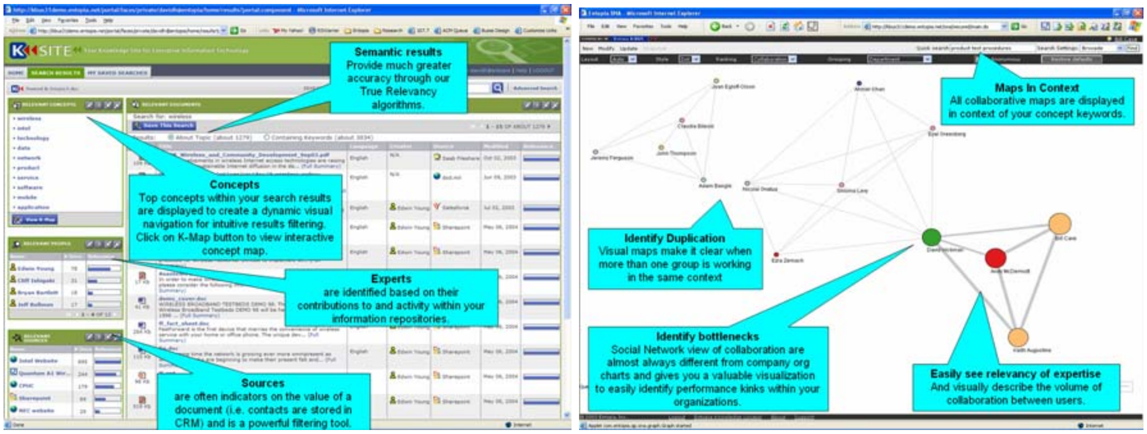
System Properties: Entopia interoperates with/has connectors for LDAP, Email (Exchange, Lotus), Documentum, web sites, file shares, and structured data sets. Unique sources require a small amount of custom work (between a couple of days and a couple of weeks). They have done this for salesforce.com but not yet for Peoplesoft or Oracle Financials. Privacy is not comprehensively addressed, however all access control list (ACL) security data is stored in metadata.

Deployments: Entopia is deployed at Saab AB (defense contractor) in Sweden (8,000 users) as well as on 1 million documents at SCA in Sweden.

Entopia states that deployment is fast with installation typically taking only a week. You simply need time to install and connect to datasources and crawl time. Subsequent crawls are incremental. No time is required to teach the systems.

Cost: $50,000 for server regardless of the number of users plus $100 per user plus a 20% maintenance fee per year.

Discussion: Following the Google IPO, Entopia couldn't get enough cash revenue so they are liquidating. While there are currently no employees, the source code is for sale and a technical support contract can be established with members of the original development team. Development was in Tel Aviv in Israel spin off from an incubator organization. There were 20 engineers 1/2 Israeli and Russian mathematician immigrants. 6 (including the Chief Technology Officer, Chief Architect, and principal product lead) spun off to create support and consulting/maintenance organization to support existing customers.



Semantic results
Provide much greater accuracy through our True Relevancy algorithms.

Concepts
Top concepts within your search results are displayed to create a dynamic visual navigation for intuitive results filtering. Click on K-Map button to view interactive concept map.

Experts
are identified based on their contributions to and activity within your information repositories.

Sources
are often indicators on the value of a document (i.e. contacts are stored in CRM) and is a powerful filtering tool.

Maps In Context
All collaborative maps are displayed in context of your concept keywords.

Identify Duplication
Visual maps make it clear when more than one group is working in the same context

Identify bottlenecks
Social Network view of collaboration are almost always different from company org charts and gives you a valuable visualization to easily identify performance kinks within your organizations.

Easily see relevancy of expertise
And visually describe the volume of collaboration between users.

Figure 12 Entopia: K-Bus Concept and Expert Search (left) and SNA (right)

8. Selecting a Vendor and Tool

Successful deployments of expert finding systems will start with choosing the right vendor and product. It is important to distinguish between features of a vendor and characteristics of a particular product. Key vendor considerations include, but are not limited to:

- *Experience* – How many years has the company been in the expert finding business? How many deployments have they done? What is their success rate? How diverse are the industry's which utilize their products?
- *Market Position* – How much of the market has the vendor captured? Are they a market leader (clear innovation, differentiation, and barriers to entry), in the middle of the pack or a laggard? Is their intellectual property protected (e.g., via a Patent or copyright)?
- *Maturity* – How mature is the software development practice of the company? Have they built methods and/or tools or support standards that enable them to more rapidly deploy and/or to mitigate common risks? Are they aware of the relative importance of different risks?
- *Skills* – Does the organization have unique competencies and talent that allow them to successfully deploy EFS? Of course this will include computer scientists, but depending upon your organization's need for consulting expertise, this could extend to other important disciplines such as human factors experts who ensure usable systems, human resources experts to help map skills, social/behavioral scientists to facilitate change management, and business process re-engineers.
- *Financial Viability* – How long has the company been in business, what is its employee size, what is its revenue history and forecast? If publicly traded, what is its debt ratio and/or asset portfolio? If available, given the importance of, well, human experts, what is the company's reputation as an employer, that is, is it a great place to work?
- *Depth and Diversity of product line* – Some vendors will provide more comprehensive offerings (e.g., search, document management, question answering, workflow) which might be integrated with their expert finding offering. But this should be balanced with the benefits that accrue when a company focuses on a particular niche.
- *Global Reach* – Does the company have a global presence? This could be a benefit for some organizations, particularly those that are geographically distributed or want access to future global innovations, but it also could pose security challenges for companies that have concerns or constraints with foreign nationals implementing solutions in their environments.
- *Breadth* – How deep is the organizations training, maintenance/support and engineering teams? How comprehensive and expensive are its training courses?

- *Reputation* – What do customers, suppliers, and competitors say about the company? Has it received awards or recognition for its product? Has it had major impact on its customers? Is the vendor honest and objective about their products features or do they hype them?

Of course it may be that a more experience, mature, and financially viable organization does not have the most innovative product, so it is important to prioritize both vendor and product considerations based on the culture and values of the implementing organization.

In addition to ensuring you have down selected to a preferred set of vendors, product considerations include:

- *Capabilities* – Perhaps the most important consideration is which of the capabilities of sources, processing, search, results display and interoperability are provided by the tool and important to your organization? These are captured in the product comparison matrix.
- *Automation* – How manual or automated is the process of expert profile creation and expert search? How important is this for your organization?
- *Performance* – How well does the system perform? In general and in your particular domain with your information sources, user classes, and business cases/workflows?
- *Product Life Cycle* – Is the product brand new or has it been through several product cycles which may imply (but not guarantee) more robust software?
- *Architecture* – Is the system flexible to the incorporation of new data sources or capabilities (e.g., social network analysis, visualization)?
- *Interoperability* – How easily can the system be integrated with and/or be made interoperable with important legacy data (e.g., personnel, financial, project) and/or systems for your enterprise (e.g., directories, document management systems, collaboration, workflow)?
- *Scalability* – Has the product been applied to organizations with the number of users (experts and expert seekers) and transactions (e.g., queries for experts), or geographic distribution that your deployment requires?
- *Security and Privacy* – Does the product provide sufficient privacy and security mechanisms to support the needs of your organization? For example, some organizations will have no problem having a system automatically process email whereas other cultures will determine this as unacceptable.
- *Time to Deploy* – How long does the product take to deploy? To how many users and with what level of capability?
- *Customizability* – Is the product customizable (ideally with pre-specified options) to your particular business needs and operating environment. Some vendors provide “skins”, or partially implemented systems for anticipated deployments such as expert finding for marketing teams, proposal teams, or human resource managers.

- *Cost* – What are the up front costs (e.g., licensing) and re-occurring costs (e.g., any transaction costs and/or annual maintenance)? How expensive and time consuming are customizations?

As with selecting a vendor, careful prioritization and assessment of these elements will increase the likelihood of selecting a product that can be successfully applied.

9. Lessons Learned for Successful Deployments

Successful deployments of expert finding systems go beyond picking the right product. First, an organization should determine if it needs to leverage commercial EF services or enable better EF within its own enterprise, or both.

The American Productivity and Quality Council (2003) identified five best practice companies who had deployed expert locator systems (Air Products, Honeywell, Intel, Northrop Grumman, and Schlumberger). Some of their more important findings included the discovery of the need to understand the work process and culture before design time, the need to iteratively develop and deploy a solution, the importance of marketing, communication, and training to increase system use, and the importance of technology. They found ongoing maintenance costs to be modest and critical need and process enablement more important than financial return as measures of success.

Based on MITRE experiences, vendor experiences, and reports such as the APQC summarized above, some key lessons learned to ensure a successful deployment include:

- *Senior Championship:* Executive leadership is important not only for obtaining (financial and human) resources, but also to help overcoming barriers (e.g., to obtain access to key data and systems to build expert profiles) as well as to motive participation in, establish realistic expectations of, and to communicate progress and successes
- *User Involvement:* Involve users (both experts and end users) from the very beginning to facilitate commitment, utility, and value. Attract them with food and promises of fame.
- *User and Culture Centered Design:* Understand the work process and culture before design time, ideally employing the skills of an ethnographer or business process expert, who can help identify opportunities and impediments to adoption early on.
- *Clear Purpose:* Tie the deployment to the enterprise mission and strategic business objectives. For what tasks will it be used (e.g., proposal team creation, troubleshooting expert discovery, project staffing, networking)? What general benefits will it provide (e.g., speed, connectivity/collaboration, knowledge/experience reuse)?
- *Measure Usage:* Communicating the value and impact of the system can start with its usage. Prepare from the very start to be able to answer questions such as: How often is the system accessed? How many distinct users are there? What kind of feedback do users provide?
- *Measure Benefit:* It is important to establish what success will be up front. Experience shows this is more likely to be enabling a key business need or process (e.g., building a proposal team, finding market intelligence) than a financial return. Nevertheless they are some key measurable benefits such as how rapidly can experts be discovered (before and after the system)? Is there a

more comprehensive discovery of experts, for example, where skills inventory is important? How do the experts like the system? Benefit measures are typically captured via surveys (on or off line, at point of use, via focus groups and so on).

- *Realism:* It may take weeks or months to determine an organization's requirements and select an expert finding tool and the same amount of time to get an initial pilot running. It is important to be realistic to all involved in terms of what is easy and what is hard, what is known and what is unknown, what is likely and what is unlikely.
- *Simplicity:* Only deploy how much capability is needed to do the job. If a simple database of names with a few keywords is all that is needed for your corporate needs, so be it.
- *Ease of Use:* Users should not need to be experts to find experts. Make finding an expert as easy as typing in a keyword or browsing a list of topics with associated experts. Locate the expert finder in an easy to find and common place (e.g., a "find" or "ask an expert" box on a corporate home page) or integrated with enterprise services (e.g., list experts at the side of an enterprise search portal).
- *Incremental Deployment:* Successful EFS have succeeded because of iterative development and deployment of a solution. In the early days of an expert finding system deployment in an organization there are many risks that need to be mitigated such as source processing, interface customization, systems integration, user adoption, training and operation. These should be piloted with a small focus group, then tested with a slight larger group, and after any necessary adjustments only then be scaled to a broader deployment.
- *Deep Understanding of Organizational Culture:* What does the organization value? What privacy and security considerations are needed? If knowledge (in this case of experts) is power, what if any informal power sources will the system challenge?
- *Privacy:* Allow users to opt-in to the system, except where there knowledge or expertise is already public (e.g., in a listserv, published document store).
- *Motivation:* Experts who are notoriously busy will not have time to fill out much less continuously maintain expertise profiles. They can be directed to do so in a autocratic organization but in other organizations they will need to be somehow incentivized (perhaps by reputation or reward or embarrassment).
- *Communications* – Marketing, communication, and training are essential and can increase awareness and adoption/use of the system.

Addressing these lessons will avoid many of the common pitfalls faced during system design and implementation.

10. Summary

This report has outlined the requirements for, challenges to, current state of the art in, and commercially available tools for expert finding. The report provides a matrix of commercial off the shelf expert finding tools, characterizes their capabilities, and discusses deployment issues. Many early adoption organizations have benefited from successful expert finding deployments. Through careful up front needs assessment, leveraging of commercial solutions, incremental deployment, and measurement of progress, any expert and knowledge intensive organizations should be able to benefit from an expert locator.

11. References

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

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

APQC. 2003. Expert Locator Systems: Finding the Answers. American Productivity and Quality Council.

Byron Dom, Iris Eiron, Alex Cozzi, Yi Zhang, Graph-based Ranking Algorithms for E-mail Expertise Analysis, DMKD03, 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, 2003.

Campbell, C., Maglio, P., Cozzi, A., and Dom, B. 2003. Expertise Identification using Email Communications. In Proceedings of ACM Conference on Information and Knowledge Management CIKM. New Orleans, LA. 528-531.

Craswell, N., P. de Vries, A, Soboroff, I. 2005. Overview of the TREC-2005 Enterprise Track. http://trec.nist.gov/pubs/trec14/papers/ENTERPRISE.OVERVIEW.pdf http://www.ins.cwi.nl/projects/trec-ent/wiki/index.php/Main_Page

Dautenhahn, K. (ed) 1998, 1999. "Socially Intelligent Agents", *Applied Artificial Intelligence*, Vol. 12 (7-8), 1998, and Vol. 13 (3) 1999.

Dautenhahn, K. (ed) 2000. *Human Cognition and Social Agent Technology*. John Benjamins.

DG. 2002. Examining the Selection Process of Choosing a Software Vendor: Expertise Location and Management Example. December 2002. Delphi Group.

Fenn, J. 1999. Skill Mining: An Emerging KM Technology. Gartner Group Report.

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

Kautz, H., Selman, B., Shah, M. March 1997. Referral Web: combining social networks and collaborative filtering. *Communications of the ACM*. 40(3): 63-65.

Kautz, B., Selman, B., & Shah, M. 1997. “The Hidden Web”, *AI Magazine*, 18(2), 27-36.

Krulwich, B. and Burkey, C. 1996. The ContactFinder agent: Answering bulletin board questions with referrals. In AAAI-96.

Lamont, J. June 2006. Finding Experts: Explicit and Implicit. *KM World* 15(6): 10-11, 24.

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.

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.

Maybury, M., D'Amore, R. and House, D. 2000a. Automated Discovery and Mapping of Expertise. In Ackerman, M., Cohen, A., Pipek, V. and Wulf, V. (eds.). Beyond Knowledge Management: Sharing Expertise, Cambridge: MIT Press.

Maybury, M. D'Amore, R. and House, D. June 2002. Awareness of Organizational Expertise. Journal of Human Computer Interaction: Special issue on "Awareness" 14(2): 199-218. http://www.mitre.org/work/tech_papers/tech_papers_00/maybury_awareness/

McDonald, D. W. Moving from Naturalistic Expertise Location to Expertise Recommendation. Dissertation thesis, University of California, Irvine, 2000.

McDonald, D. W. and Ackerman, M. S. 2000. Expertise Recommender: A Flexible Recommendation System and Architecture. CSCW. December 2-6, Philadelphia, PA, 231-240.

Mowbray, B. Challenges in Locating Experts. American Productivity and Quality Council.

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

Swartz, M. F. and Wood, D. C. M. 1993. Discovering shared interests using graph analysis. *Communications of the ACM*. 36(8): 78-89.

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.

Zhu, J., Gonçalves, A.L., Uren, V.S., Motta, E., Pacheco, R. 2005. Mining Web Data for Competency Management. In Proceedings of Web Intelligence 2005 (WI'2005), France, September 19-22, pp. 94-100.

Dumais, S. and Nielsen, J. 1992. Automating the Assignment of Submitted Manuscripts to Reviewers. SIGIR 1992: 233-244

Yimam-Seid & Kobsa, Expert Finding Systems for Organizations: Problem and Domain Analysis and the DEMOIR Approach. Journal of Organizational Computing & Electronic Commerce 13 (1), 2003.

12. Acronyms

ACL	Access Control List
ASP	Application Service Provider
CoP	Community of Practice
COTS	Commercial Off-the-shelf
CSCW	Computer Supported Cooperative Work
DB	Database
D&D	Denial and Deception or the Intentional Suppression or Distortion of the Truth to Mislead Others
EFS	Expert Finding System
ELS	Expert Location System
ERP	Enterprise Resource Planning
FAQs	Frequently Ask Questions
IE	Information Extraction
KB	Knowledge Base
LDAP	Lightweight Directory Access Protocol
MAP	Mean Average Precision
NE	Named Entity (e.g., people, organization, location, facility)
ODBC	Open DataBase Connectivity (A standard database access method developed by the SQL Access group in 1992)
OLAP	On Line Analytical Processing
ROI	Return on Investment
TREC	Text Retrieval and Evaluation Conference
W3C	World Wide Web Consortia

13. Terminology Definitions

ASP An Application Service Provider (ASP) is a business model that provides computer-based services that it owns and operates over the network.

Context The circumstances, conditions, or environment that surround something such as the time, space, visual, auditory, linguistic, task, historical, political, or religious context.

Culture The values (beliefs, ideas, opinions), norms (expected behaviors), and artifacts (e.g., religious, culinary, artistic, clothing) that characterize a group of humans.

Expert An individual with authoritative knowledge, skill, or experience in a particular domain or subject matter.

Expertise An expert's area of special knowledge or skill.

EFS An Expert Finding System (EFS) enables users to discover domain or subject matter experts in order to hire or acquire their knowledge.

ERP Enterprise Resource Planning (ERP) systems typically integrate a broad range of data and processes from an organization including financials, human resources, supply chain, customer relationship management, warehouse management, and so on. Examples of ERP systems include SAP, Oracle, and PeopleSoft.

Information Extraction Automated detection and tagging of named entities (e.g., people, places, things) in unstructured text.

Massive Data Petabytes or exabytes of data (e.g., web, geospatial, imagery, call data).

MAP Mean Average Precision (MAP) is the mean of the average precisions over a set of queries after each document is retrieved. This measure gives better scores to techniques that return more relevant documents earlier.

Media Text, image, audio, and video files in a broad range of formats.

Modality Different human sensory modalities such as auditory, visual, and tactile.

Multilingual Multiple human languages such as English, Arabic, or Chinese. The data are in many natural languages, not just English.

Multiscale Analysts need to deal with individual documents, collections of documents, and even multiple repositories.

OLAP On Line Analytical Processing (OLAP) supports the analysis of structured data in order to perform business intelligence.

Recommender A system that takes previous user selections and attempts to predict what.

System Kinds of items or resources will most likely satisfy them in the future (e.g., Amazon.com recommends books based on past purchases).

Sociology The study of individuals, groups, and societies including their identity, relationships, and interactions.

Streaming Data not at rest on a disk, but that are distributed in real time (e.g., streaming video).