Discovering Distributed Expertise

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

Abstract

This article provides an overview of expert finding systems which assist in the discovery of distributed expertise. We summarize requirements for expert finders, the current state of the art, and our corporately deployed expert finding system and its performance. We describe research directions in distributed expertise discovery and visualization.

Introduction

In a global economy, expertise is increasingly likely to be created and operate in a highly distributed fashion (Friedman, 2005). While clusters of competence will emerge (Porter 1998), the ability to find and foster global communities of interest and put diverse skills together for competitive advantage will be a critical success factor in a global market. Automating this competence discovery process will be come a key requirement.

Expert Finding Systems (EFS), also called Expertise Location Systems (ELS) enable users to rapidly, accurately, inexpensively and securely/privately discover experts, expert networks, and the knowledge they contain. 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.

Challenges

Expert finding is a difficult task because experts and their skills and knowledge are rare, expensive, (unevenly) distributed, difficult to qualify, varying in level, continuously changing, and often culturally isolated and oversubscribed.

To complicate this, expert seekers typically have poorly articulated requirements, have no insight into past performance of experts, and often have difficulty judging a good expert from a bad one. Finally, users' complex problems often require the combined wisdom of multiple experts.

Commercial Systems

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. Table 1 compares products in terms of weather they provide full, partial or no support for key functions such as:

- *Sources* processed to determine expertise (e.g., expert self declarations, communications such as email, publications such as documents, behavior such as searches)
- *Processing* performed systems (e.g., ranking experts on the basis of entities extracted from their publications, social network analysis, foreign language processing)
- *Searching* facilities supported (e.g., keyword, Boolean, natural language query, taxonomic browsing).
- *Results* presentation (e.g., ordered lists of experts, lists of documents produced or used by experts, concepts related to the topic of expertise)
- *System properties* such of the degree of interoperability, privacy, and operational deployment.

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.

Performance

While there are no performance benchmarks of commercial systems, 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 (http://trec.nist.gov/). 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%.

Submitted to AAAI Fall Symposium Regarding the "Intelligence" in Distributed Intelligent Systems (RIDIS), Arlington, VA

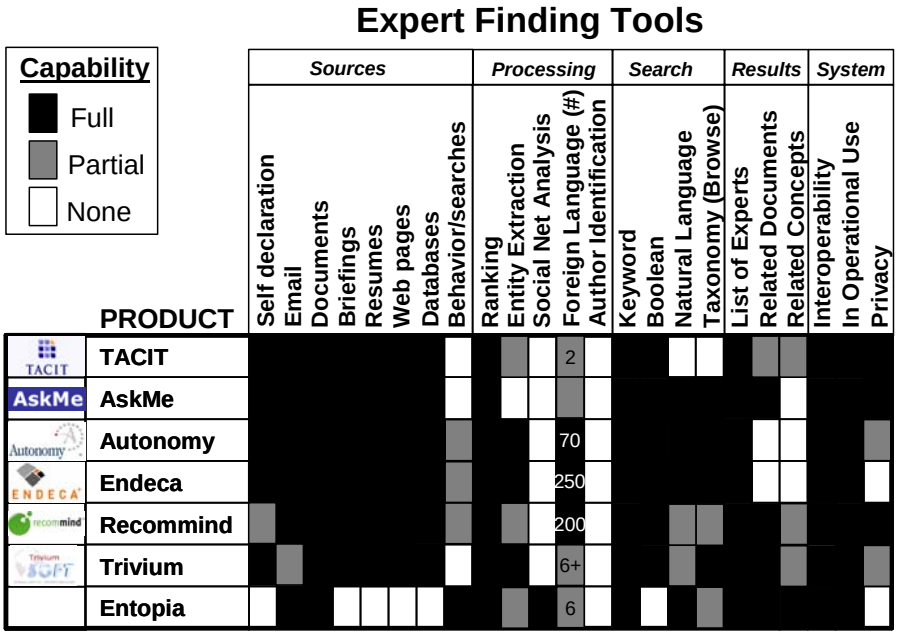

Figure 1. Commercial Expert Finding Systems

Deployment Lessons Learned

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. For a complete detailed report of EF systems, their specific features, costs, and deployment guidelines, see Maybury (2006).

MITRE Expert Finder

MITRE's Expert Finder (Mattox, Smith, and Seligman 1998) was aimed at placing a user within one phone call of an expert based on a simple user query. Given a query, the system ranked 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 had published in their resume or document folder (a shared, indexed information space). Integrated with MITRE's corporate employee database, employees were ranked by frequency of mentions, pointing to sources in which they appear. In empirical evaluations, in spite of the fact that human agreement regarding expertise was surprisingly low (60% or less), over 40% of the experts returned by Expert Finder were judged by humans as experts (a measure of "precision"). Expert Finder also found about 30% of all the experts identified by human experts (a measure of "recall") (Mattox, Maybury and Morey, 1999).

Figure 2. MITRE's Operational Expert Finder

Figure 2 and 3 illustrate the operational system that MITRE deployed corporately based upon experience with the initial prototype. MITRE leveraged a corporate deployment of Google to index content such as employee publications and project descriptions. Users search using a simple keyword interface shown in 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 enables validation of expertise as well as ac-

cess to the expert's artifacts which are often of interest to the expertise seeker. 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 screen shot shown in Figure 3. This 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 in a corporation of over 6000 employees.

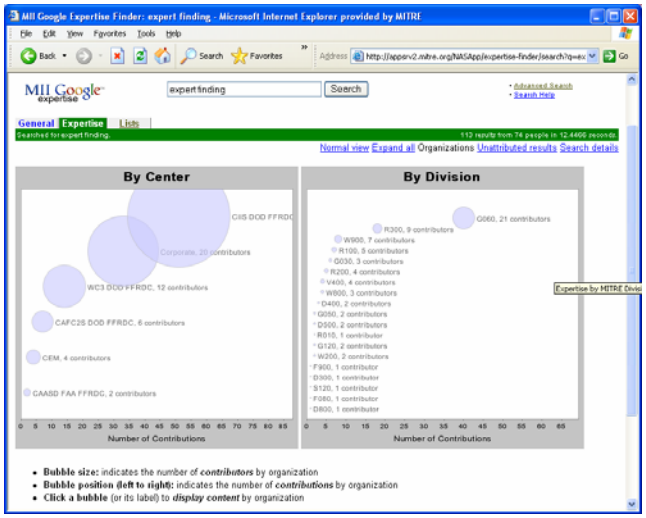

Figure 3. Expertise Distribution by Organization

Visualizing Expertise

As illustrated in Figure 2 and 3, visualizing expertise relevancy in a search results like list form as well as visualization the distribution of expertise across organizational boundaries are important facilities for speeding discovery and enhancing awareness of expertise. Another noteworthy visualization of skills is that offered by the commercial product Trivium. Trivium interoperates with enterprise resource planning systems (e.g., Peoplesoft) to build skill maps and can be used in both a reactive mode (e.g., what skills, experiences, interests does an organization have) as well as a tool to assess organizational skills (identifying primary, secondary, and weak skills). The tool's Capability Tree map of skills (exemplified in Figure 4) 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 human resources diagnostic project, including a skills collection survey of 500-1500 people, takes on average 4-6 weeks to complete. At the end of the survey users can create a Capability Tree and use the application for on-going skills management. The example Trivium Tree Map shown in Figure 4 displays employee skill frequency and relationships. Color is used to encode skill level. The darker hued colors at the trunk of the skill tree indicate more frequent skills, those ligher hued ones at the top indicate rarer skills. Each building block in the tree repesentes a key skill set. Skills on the left of the display are ranked according to decreasing frequency and employees listed on the right are ranked on the basis of match with a selected skill.

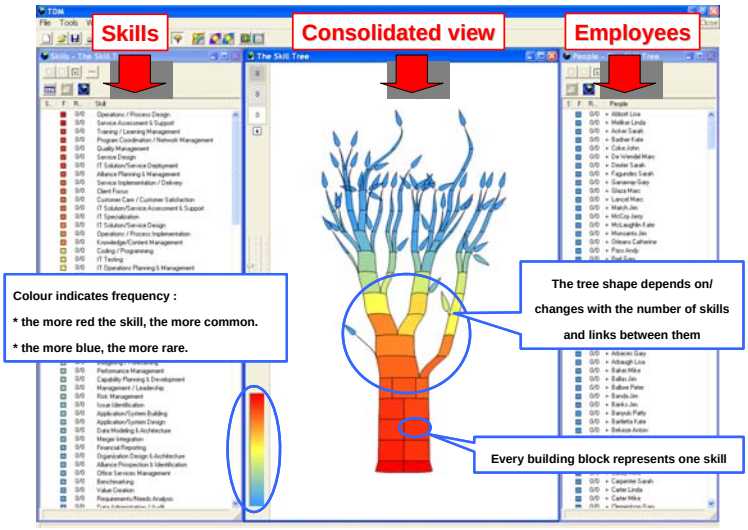

Figure 4. Trivium Tree Map

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 the 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 an this tool.

Evaluation

While we can create a system to identify and visualize expertise distributed geographically and organizationally across a corporation, how do we assess how well it is working? As describe above, our measure of accuracy for the MITRE Expert Finder was based on precision and recall, measures inspired by research in information retrieval. Related, the measure used by the TREC evaluations is Mean Average Precision (MAP). In addition to technical measures, however, other measures of merit can be more important for an organization. For example, benefits beyond speed and quality of retrieval might include:

- *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.
- *Expert Disclosure*: Does the appearance of expert finding services encourage experts to publish expert profiles or their expert content?

While our expert finder performs its searches in seconds, measuring these other effects remains an area for future research.

Future Research

There are a number of remaining challenges that are important areas for future research. These include:

- *Evidence*: With diverse sources and indicators of expertise, how do you assess the provenance and quality of that evidence? Should it change based upon the query and/or purpose for the expertise?
- *Validation*: How do we assess the level and range of expertise of an individual? Do we rely on textual artifacts, human feedback, social network analyses?
- *Visualization*: What is the best method to display individual and/or group expertise? How can we effectively display communities of experts and their relationships?
- *Privacy*: How do we maintain the privacy and control accessibility to experts while at the same time enabling effective access for those who have a legitimate need to contact the expert? For example, mechanism such as automated de-identification (Wellner et al. submitted) could enable privacy preserving expertise discovery.

Expertise management and expert discovery promises to become an increasingly valuable and accessible capability across academia, industry, and government.

References

1. Friedman, T. 2005. *The World is Flat*. Farrar, Straus and Giroux.
2. 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.
3. 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.
4. 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.
5. Maybury, M. July 2006. Expert Finding Systems. MTR 06B-40.
6. Porter, M. 1998. *The Competitive Advantage of Nations*. Free Press.
7. Wellner, B. et al. "Rapidly Retargetable Approaches to De-identification in Medical Records". submitted to *Journal of the American Medical Informatics Association*.

Acknowledgements

We thank Lisa Costa and David Poutre for their support of evaluation of commercial expert finding tools.

Biography

As Executive Director of MITRE's Information Technology Division, Dr. Mark Maybury is responsible for the direction of MITRE advanced research and development for intelligence and defense systems. Mark has organised international conferences, given tutorials, and published over fifty articles in the area of language generation, multimedia presentation, text summarization, intelligent information retrieval and analysis. Mark is editor of *Intelligent Multimedia Interfaces* (AAAI/MIT Press, 1993), *Intelligent Multimedia Information Retrieval* (AAAI/MIT Press, 1997),. *New Directions in Question Answering* (AAAI/MIT Press, 2004*)*, co-editor of *Readings on Intelligent User Interfaces* (Morgan Kaufmann Press, 1998), *Advances in Text Summarization* (MIT Press, 1999) and *Advances in Knowledge Management: Classic and Contemporary Works* (MIT Press, 2001) and co-author of *Information Storage and Retrieval: Theory and Implementation. 2nd Edition (*Kluwer Academic, 2000) and co-editor of *Knowledge Management* (MIT Press 2000). Mark serves on the Board of the Object Management Group.

Keywords

Expert finding, expertise location, expertise management, expertise visualization, expertise validation