Lotus

White Paper

Locating Organizational Expertise with the Lotus Discovery Server

January 2002

A Lotus Software White Paper

IBM

© Copyright 2002 Lotus Software
IBM Software Group
One Rogers Street
Cambridge, MA 02142

Contents

Introduction 1
- Automating Expertise Location 1
- The Lotus Discovery Server 2

Linking Knowledge Resources to Expertise 3

How the Discovery Server Identifies Expertise 7
- Setting Up Directory Synchronization 8
- Scheduling Content Spiders 13

Locating Organizational Expertise 16
- Profiling Interactions with Content 16
- Managing Individual Profile Documents 18
- Knowledge Audits 20

Conclusion 21

For More Information 22

Introduction

With the dramatic growth of the Internet, intranets, e-mail and collaborative applications, vast amounts of information are available to each of us. Customer records, order information, support call records, discussion forums, internal research memoranda, industry publications, newswires, collaborative project tracking data, trade journals, policy and procedure databases: these are just a few of the nearly innumerable types of information that surround us — many of which change daily.

In today's competitive business environment, trends such as rising customer expectations and compressed product development cycles are rapidly forcing organizations to look beyond traditional methods of capturing, storing and distributing data. Effective management of organizational knowledge resources has become mission critical.

No matter how much information is available, however, effective decisions frequently require the input of experienced people. It is by requesting guidance, posing options and group-thinking strategies that we benefit from what others have learned in related situations. And asking someone who knows is frequently the fastest way to find the correct answer to a question.

Because the expertise of others is so immensely valuable to the decision-making process, people invest considerable effort in trying to find the "right person" (within or outside their own organization) to consult on a problem. Often, anecdotal information about who knows what is passed down informal channels from contact to contact. Sometimes the recommended person is available; sometimes not. Perhaps they left the company, for example, or have changed jobs. In short, we frequently need to know precisely who are the *current* experts on specific subjects.

Automating Expertise Location

The ability to systematically locate current expertise resources, across different divisions and physical locations, is of significant competitive value to organizations. Employees, partners and customers get faster, better answers to questions leading to better decisions, improved responsiveness and customer service, reduced cycle times and ultimately competitive advantage. But time, distance and employee turnover can make it difficult to reliably locate experts by ad hoc methods. Automated systems can speed up the process, but many are only as good as the information employees do (or do not) provide about themselves. Experience has shown again and again that users do not keep personal information up-to-date when required to do so manually. Expertise location systems must automate not only the retrieval of information about expertise, but also the gathering of information about expertise.

The Lotus Discovery Server

The Lotus® Discovery Server™ addresses enterprise requirements for efficiently and systematically locating current expertise resources, by automating expertise profiling. The Discovery Server finds and organizes the dynamic relationships among information, people and the activities of people to automatically pinpoint expertise.

The Discovery Server is a key component of content managent and search solutions that seek to organize relevant content and expertise to address specific business tasks and processes. Discovery Server plays an important role in the overall Lotus and IBM® knowledge management strategy, which provides collaborative e-business solutions that bridge people and knowledge to optimize an organization's business transactions.

Linking Knowledge Resources to Expertise

The Discovery Server "crawls" knowledge resources in order to evaluate their content, using content *spiders*. Spiders are currently available for a wide range of data sources, including:

- Web and intranet sites
- Web discussion forums
- File systems (filters enable full text search of documents created with business productivity tools like Microsoft Office and Lotus SmartSuite®, .pdf, etc.)
- Back-end systems
- Lotus Domino™ applications
- Lotus Domino.Doc™ file cabinets
- Lotus QuickPlace™ team workspaces

The Discovery Server extracts and securely stores a wide variety of information about the content it finds. To link people to their expertise, the Discovery Server examines the activities employees perform, including the topics of documents they've read and created, discussion participation, the contents of e-mail messages, and much more. This ongoing, dynamic process is integral to the creation of both of the Discovery Server's primary outputs:

- The Knowledge Map, or K-map interface, a complete, hierarchical taxonomy of an organization's machine readable knowledge resources. The K-map can be either searched or browsed, to locate both information and expertise.
- Expertise Profiles, which document each employee's affinities, skills, interests, project experience, education and job type.

End users browse and search the K-map taxonomy categories, content, and Expertise Profiles via Web browsers or Lotus Notes® clients, as shown in Figures 1 and 2. Custom applications can also leverage this Expertise Locator service; the K-map is URL addressable, and thus be incorporated into other applications. In addition, the Discovery Server includes a set of Java-based APIs that can be used to integrate and extend the organizational knowledge discovery environment. For example, using the Discovery Server API Toolkit and sample Web pages provided it's very straightforward to incorporate Discovery Server managed content and expertise data into other application or portal interfaces.

As Figure 1 illustrates, the Discovery K-map interface presents categorized and ranked content in relation to search criteria and topic categories, enabling users to efficiently identify the information and people that will be most helpful.

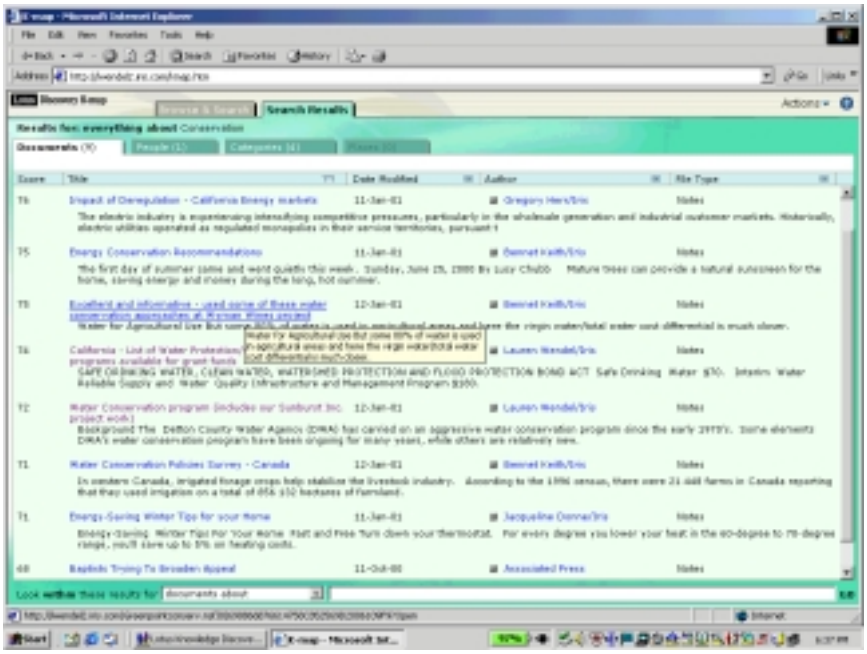

Figure 1: The Knowledge Map makes it easy to find the right information and expertise.

The Discovery Server also makes it easy to connect with experts. For example, clicking on a person's name in the K-map brings up his or her Expertise Profile, as shown in Figure 2. The information seeker then has the option to automatically address an e-mail to the person, initiate an instant message to them, or add their name to a community or shared place within a portal application such as WebSphere® Portal Server/K-station or other portal server applications.

Figure 2: Discovery Expertise Profiles identify contact information for subject matter experts.

Discovery Expertise Profile and Category and document information can be incorporated into any application, by simply embedding the Discovery Server K-map URL into an application or by using the Discovery Server Java APIs to customize incorporation into any application or portal infrastructure. An example, showing integration with a WebSphere Portal Server application is shown below:

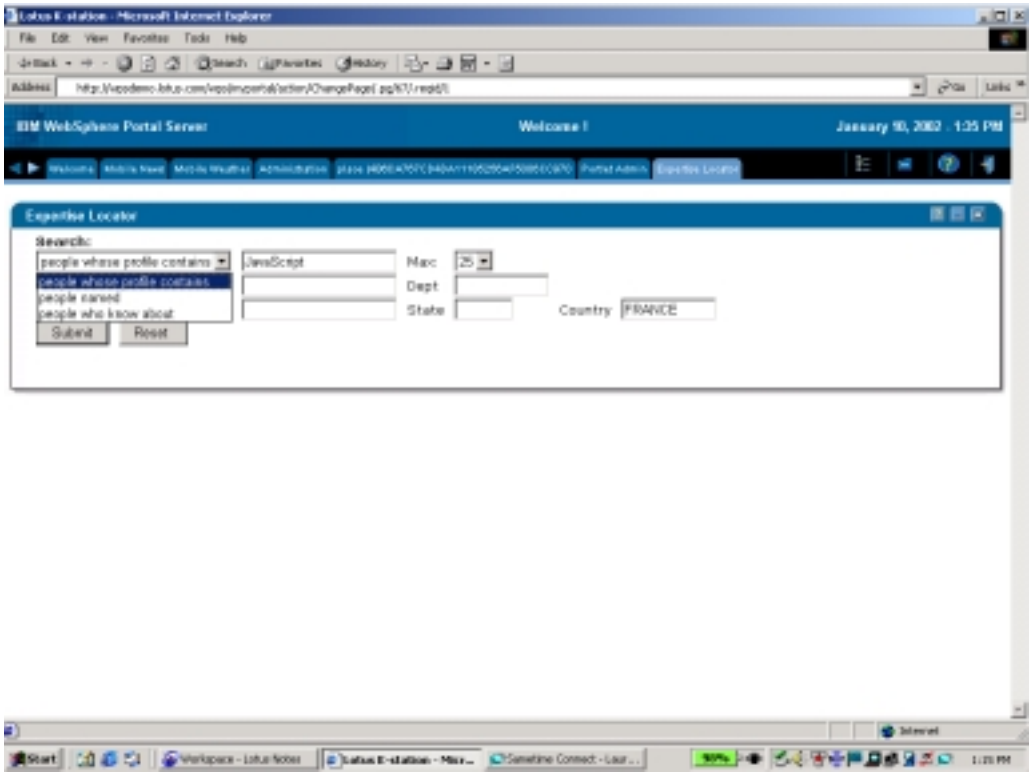

Figure 3: Searching Discovery Server Expertise information from a WebSphere Portal Server application.

Users can edit their own Expertise Profiles, as shown in Figure 4.

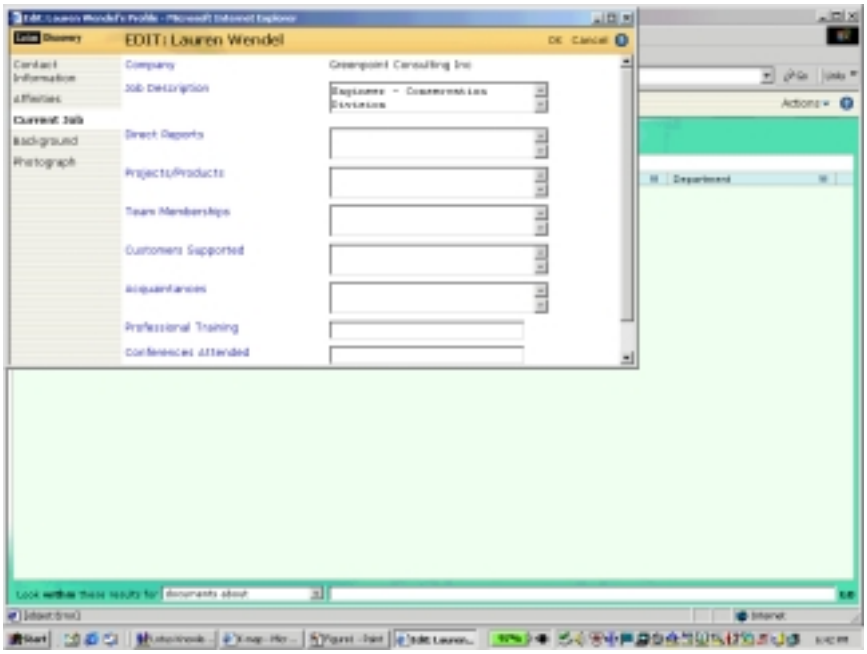

Figure 4: Editing an Expertise Profile.

The Discovery Server can also generate Knowledge Audits, which show what organizational data is most frequently accessed, and hence potentially of greatest strategic value. Usage trends are tracked over time across all repositories, enabling knowledge officers and line-of-business managers to make informed decisions related to the utility of information.

The remainder of this white paper explains how the Discovery Server identifies, analyzes, categorizes and reveals organizational expertise.

How the Discovery Server Identifies Expertise

The Discovery Server obtains public user data from the Domino Directory or other LDAP server directory. It combines this contact data (telephone number, work location, e-mail address, etc.) with the user's contributions and activities to build an Expertise Profile. Usage activity metrics, which the Discovery Server tracks, determine the values that pertain to the user across different categories in the K-map. Figure 5 illustrates these processes.

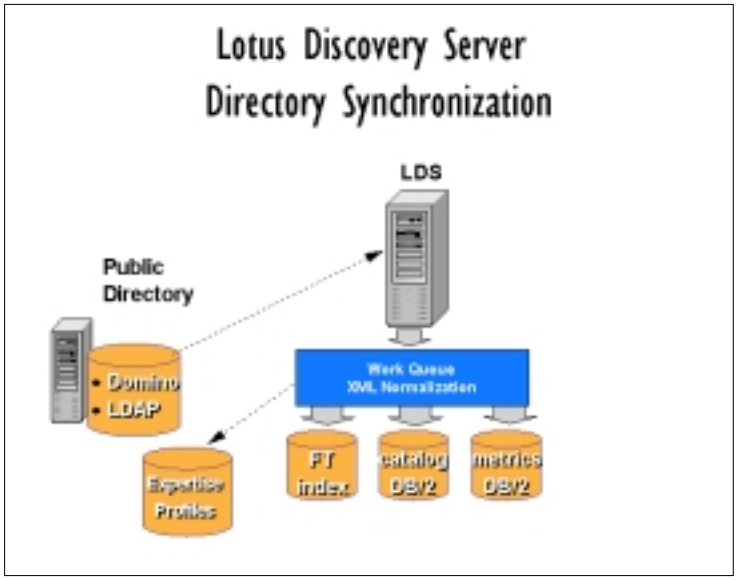

Figure 5: Creating Expertise Profiles from Directory Source Data.

Because users rarely find time to update their own profiles, the Discovery Server generates and updates them automatically, both by synchronizing with public directory sources and by discovering employees' evolving expertise and affinities. This automated capability greatly benefits both system administrators and end users. Users have the ability to modify their personal profiles with background data, to accept or reject proposed expertise indicators, and/or to accept and publish a profile for others to access.

The Discovery Server synchronizes information from public directory sources on a scheduled basis. As shown in Figure 5 above, the Discovery Server uses this information to create initial profile records. As the Discovery Server aggregates information about the activity of each person profiled, it factors it into that person's Expertise Profile document (all of which are stored in the Discovery Server profile database PEOPLE.NSF).

Setting Up Directory Synchronization

To initiate directory synchronization, system administrators identify public directory sources of people information using the Discovery Server Administrator.

As shown in Figure 6, the Administrator application provides options that let administrators specify both primary (authoritative) and secondary sources of information about people. The primary directory source is used to initially create users' Expertise Profiles. Because organizations frequently store additional information about people in skills databases, HR applications or project teamrooms, the Discovery Server can access this information and store it automatically as supplemental information to user's profile forms.

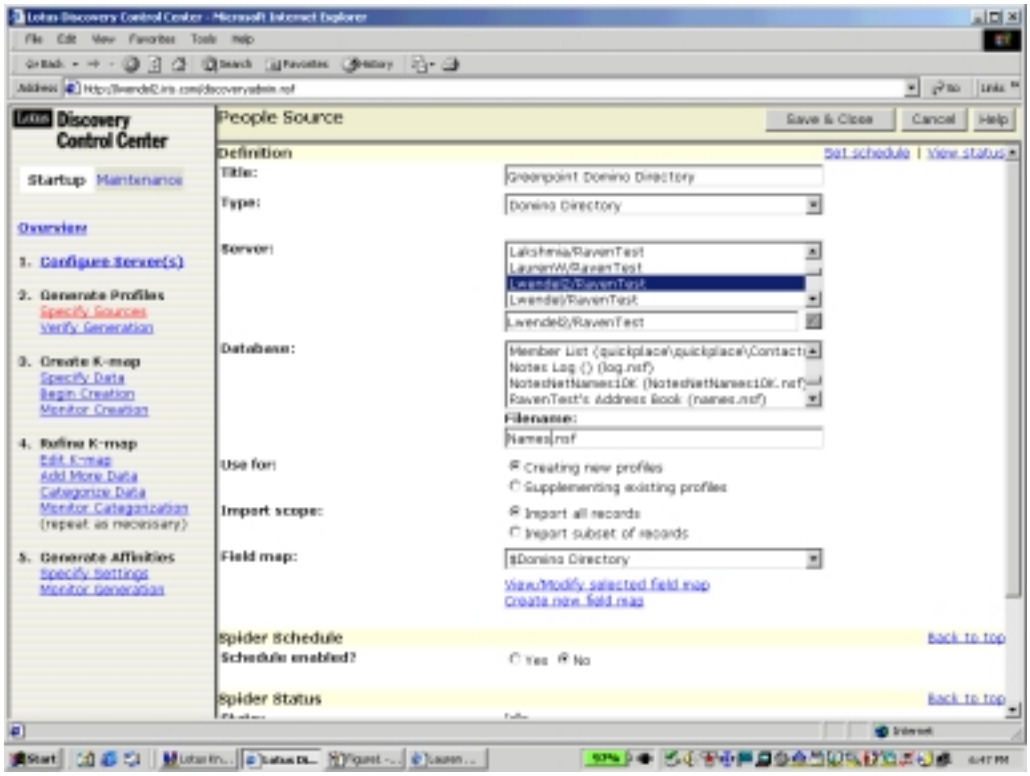

Figure 6: Specifying sources of information about people.

Figure 7 shows how administrators can specify fields in external directory sources for the Discovery Server to match to its profile form fields.

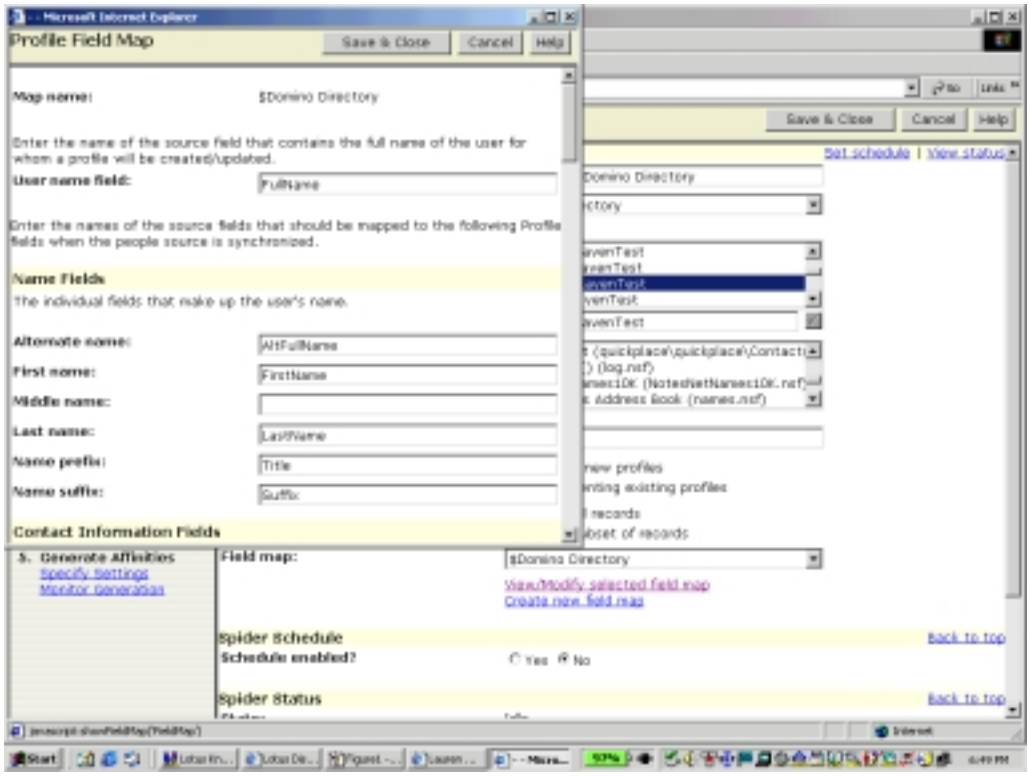

Figure 7: Mapping fields in external directories.

Using the Discovery Server Administrator, the Profile Source Spider can be scheduled to access the external directory used as the source data to create profiles within the Discovery Server directory. These profiles determine those individuals in the organization whose activity is monitored for expertise location purposes.

As Figure 8 illustrates, the system administrator should enable the Profile Source Spider, Metrics and Profile Updating tasks before scheduling the external directory for synchronization with Discovery Server.

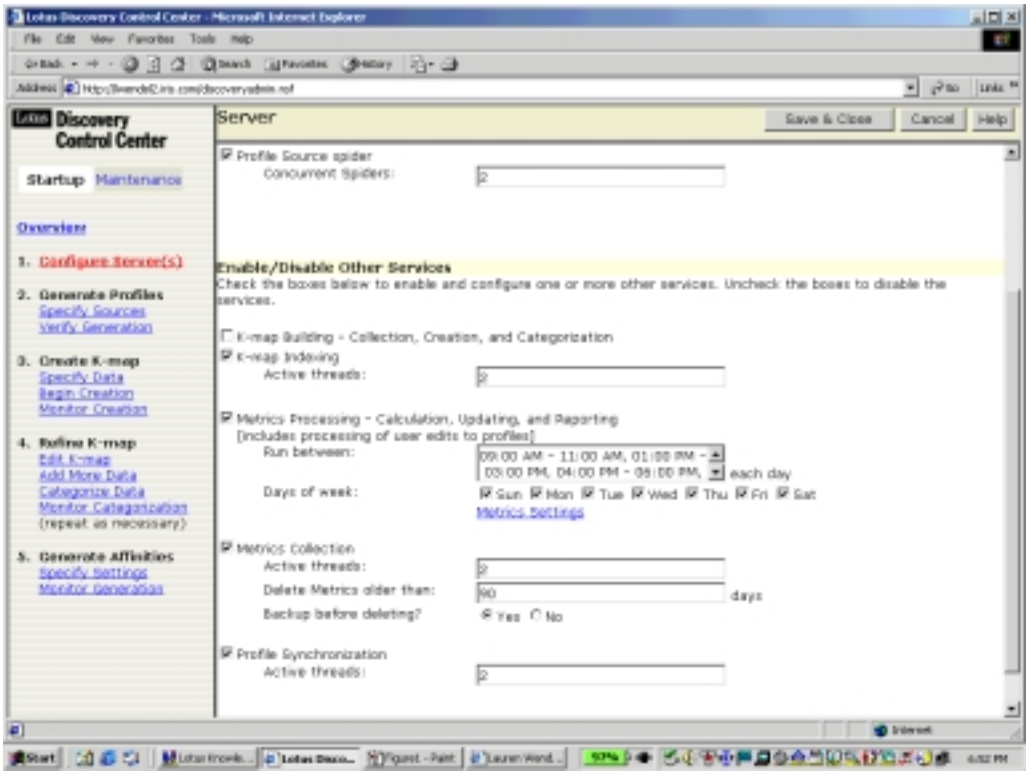

Figure 8: The Discovery Server Task List.

Typically, directory repositories are scheduled to be synchronized once per day with the Discovery Server, although it is possible to schedule more (or less) frequent updates. Figure 9 illustrates this process.

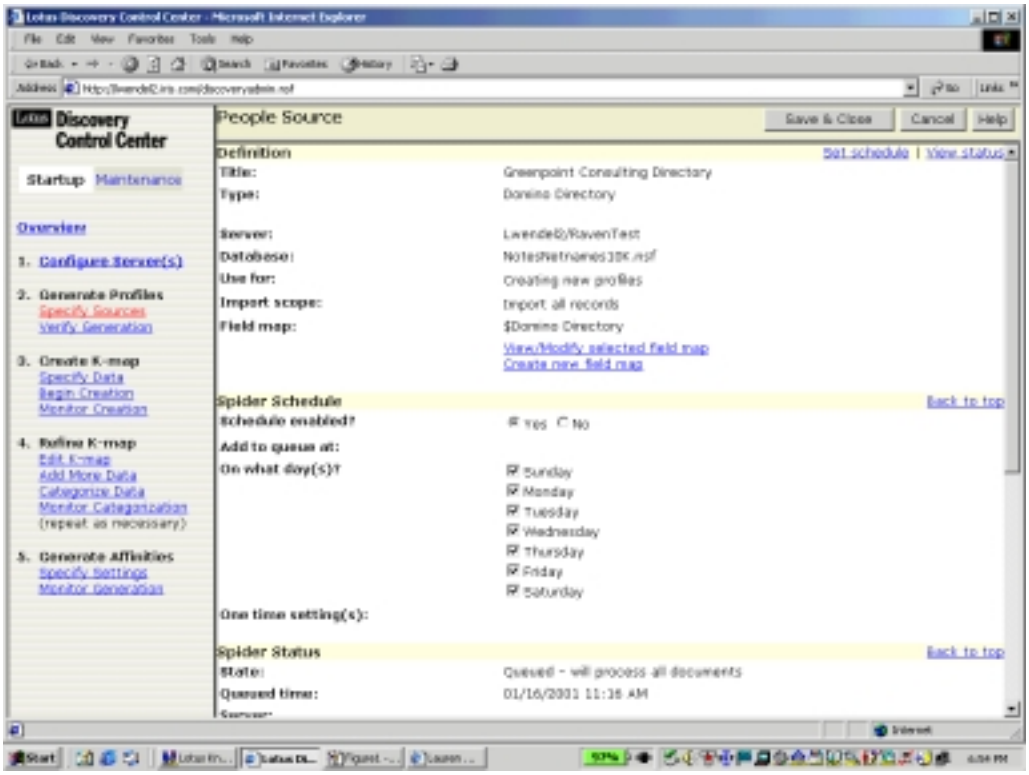

Figure 9: Scheduling profile synchronization.

Once Profiles are created within the Discovery Server, it is possible to synchronize additional, supplemental information, such as skills inventory lists maintained in other applications, into users' Expertise Profiles. An example is shown in Figure 10.

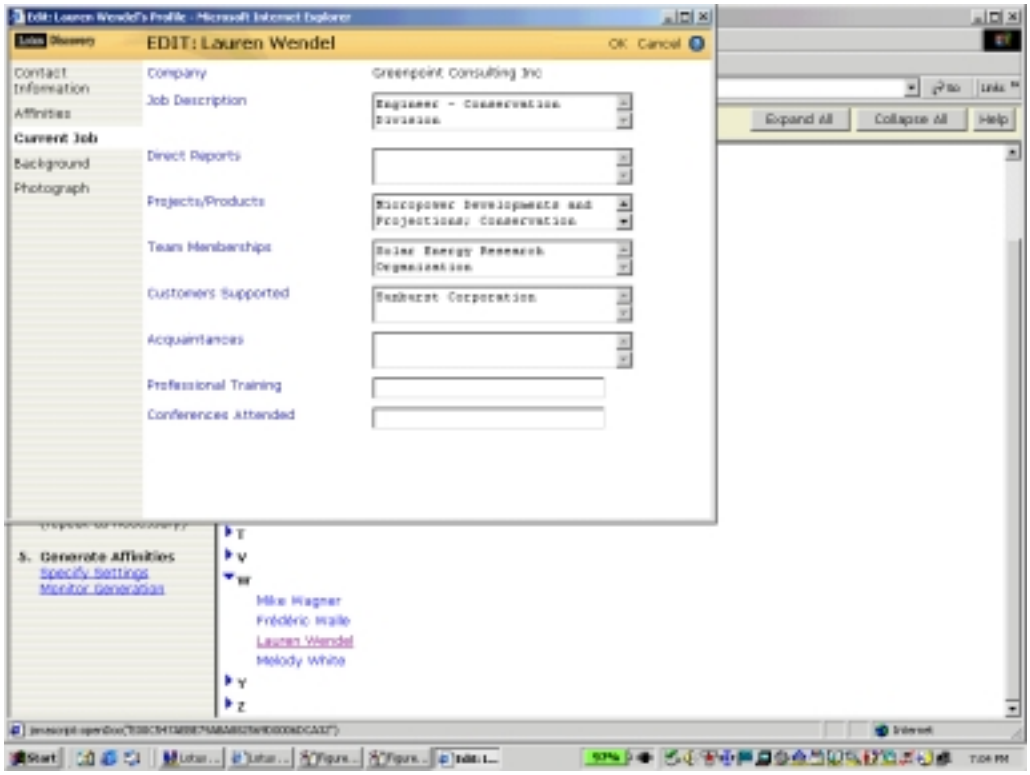

Figure 10: Synchronizing supplemental skills information.

Scheduling Content Spiders

Since different types of content need to be scanned more or less frequently, the Discovery Server allows administrators to schedule how often content spiders will crawl various content repositories. The knowledge resources found in these repositories are automatically incorporated into new or existing K-map categories. Figure 11 illustrates this process.

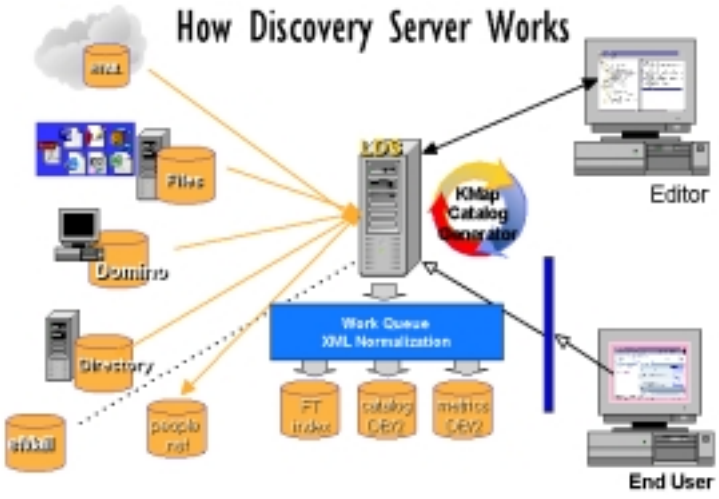

Figure 11: How the Discovery Server spiders content.

As the spiders access knowledge repositories, information such as source document location, access controls, usage activity, authorship, title and subject data are returned to the Discovery Server. The server passes the incoming data to several work queues, which process it in parallel. The indexing of new content for full text search, the storing of relevant data within the metrics database, and additional processes serve to analyze this data, using it to cluster related documents to categories, calculate document relevance values and relate user affinities to content categories. Discovery Server processes are multithreaded, and may be installed across server platforms for enhanced scalability.

As it builds or refreshes the K-map, the Discovery Server organizes related documents into category areas, across all spidered repositories. The K-map presents this organizational scheme as a hierarchical, readily navigable index of terms, which displays relationships evaluated by the Discovery Server, and their degree of relationship (or relevance) to the topic.

Figure 12 displays the top-level category list of a K-map generated through analysis of several repositories that contain marketing reports, product support data, discussion databases, project teamrooms and retail channel announcements.

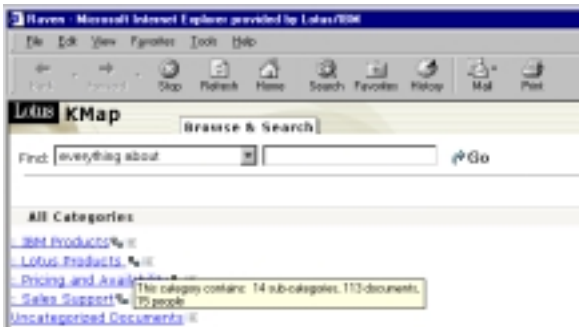

Figure 12: Top-level categories of a sample K-map.

The Discovery Server includes a K-map Editor, designed to be used to refine the organization of content and to enhance the usefulness and relevance of topic labels and other terms within the organizational taxonomy. In particular, individuals with appropriate edit privileges can revise the category labels that the Discovery Server generates automatically, and reassign documents to different areas of the K-map. This "post-processing" serves to customize the K-map to suit the organizational culture, thus optimizing its value for end users.

As shown in Figure 13, the K-map Editor displays the documents that make up each category, so the K-map can be refined down to the level of individual documents.

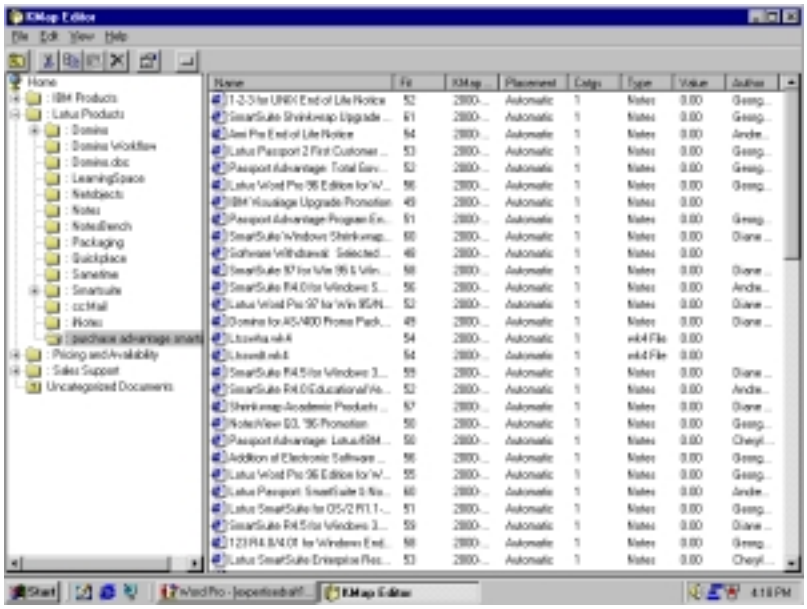

Figure 13: The K-map Editor.

As human editors refine the K-map over time, the Discovery Server tracks all modifications made to category labels and the organization of documents within categories. These changes are “remembered” and reapplied each time the K-map is automatically refreshed, so that the document classification processes can be continually improved. In this way the K-map can be “trained” in accordance with the evolving needs and preferences of the organization.

Locating Organizational Expertise

As touched on above, the Discovery Server automatically relates the activities of profiled employees to the content and category areas of the K-map. As individuals interact with the content found in various repositories, such as documents, discussions, team workspaces, etc. that comprise a category, the Discovery Server assesses the quality, frequency and recency of those interactions to calculate the strength of each person's affinity, or expertise ranking, within that category. Actions such as authoring a document, responding to a discussion topic, editing a document and creating a link to a document are among the actions that the Discovery Server monitors. It updates individual metrics records based on that detail, and then aggregates the result to calculate specific expertise rankings.

Using this metrics data, the Discovery Server automatically calculates document values and expertise ratings. System administrators determine a threshold setting to be applied when evaluating affinity relationships. Typically, a setting of 50% to 60% is a reasonable threshold to apply when determining the strength of affinity to category areas (keeping in mind that employees make the final decision about the validity of an affinity). Affinities below that threshold are deemed to be "weak" and are not presented to users.

Profiling Interactions with Content

The Discovery Server runs specific processes, referred to as metrics calculations, which act on the data gathered through various metrics collection tasks, in order to calculate the relative value of all the content it tracks. Each element of content stored and monitored within the metrics data store is called an "entity." Among the entities the Discovery Server stores and monitors are: Person, Document, Term, Topic, Place (i.e., an object type to track community place types such as Websphere Portal Server/K-station or other portal server Community Place), Database and Server. The Discovery Server creates a record and assigns an ID, which are used to identify each entity and the actions it performs over time. Figure 14 lists representative metrics.

Entity 1	*Metric*	*Entity 2*	*Value Represents*
Person	is author of	Document	Documents authored
Person	creates links to	Document	Links created
Person	responds to	Document	Responses to document
Document	contains links	n/a	Links in document
Document	has responses	n/a	Responses to documents
Document	has recent accesses	n/a	Number of recent accesses
Document	is linked to	Document	Number of links from document to document
Document	is response to	Document	Document is response to second document
Category	is referenced in	Document	Count of references to category
Category	affinity	Person	Computed affinity of person for this category

Figure 14: Sample Discovery Server Metrics.

System administrators can modify the weighting, and hence relative affinity valuation, applied to various Discovery Server metrics, as Figure 15 shows. For example, the Discovery Server attributes a higher weighting by default to actions such as authoring and editing documents than it does to reading documents via the K-map. These default valuations can be altered based on organizational preferences. For example, responding to a document may be determined to have greater contribution value than editing a document, and can be given a correspondingly greater weight.

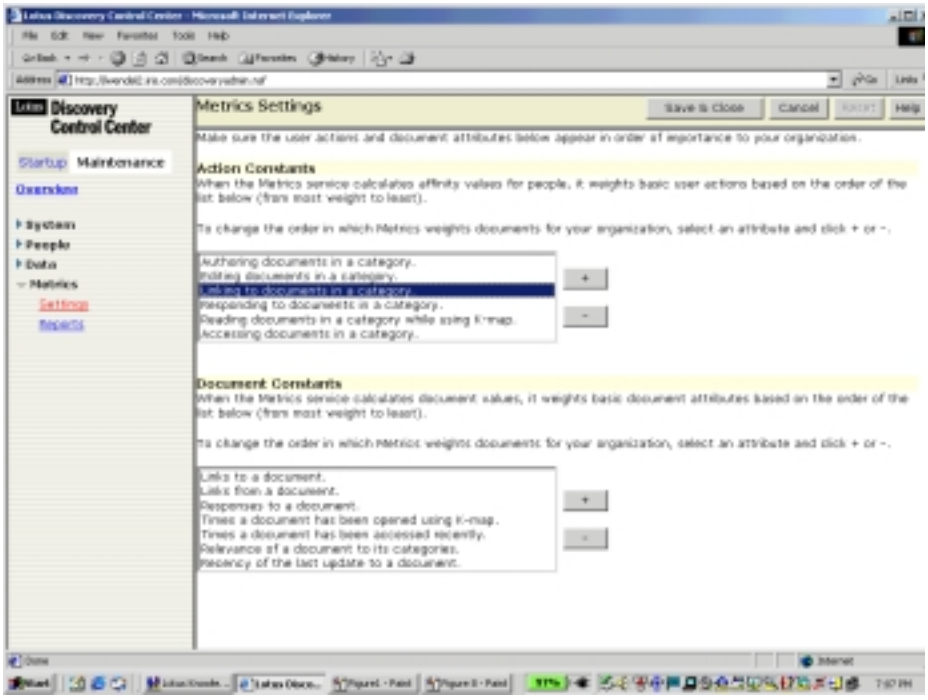

Figure 15: Customizing the relative weight of expertise and document value metrics.

The Discovery Server also provides options to modify the default metrics applied to calculate document values, as shown in Figure 15.

Document value rankings are also factored into the evaluation of peoples' affinities. Individuals who have authored documents that are useful to others, for example, as calculated by the Discovery Server metrics processes, will have a correspondingly greater affinity to the relevant category areas.

Managing Individual Profile Documents

The person profiled in an Expertise Profile is the owner of that profile, and by default has the access permission necessary to read and edit his or her own profile information. This enables employees to maintain control over information about them that is made available to others. Profile owners can, for instance, add additional background data to the profile, including previous work experience, skills and industry associations, education, interests, photos and so forth.

Automatic Notification of Affinity Changes

As described above, the Discovery Server uses the metrics information it gathers to continually recalculate the affinity strengths of each employee relative to others profiled in the organization. When the Discovery Server calculates an affinity greater than the specified threshold, it sends an e-mail (see Figure 16) to the profiled user notifying them that a strong affinity to a K-map category has been calculated. The message asks the user to evaluate the assessment, and approve/disapprove the change to their Expertise Profile.

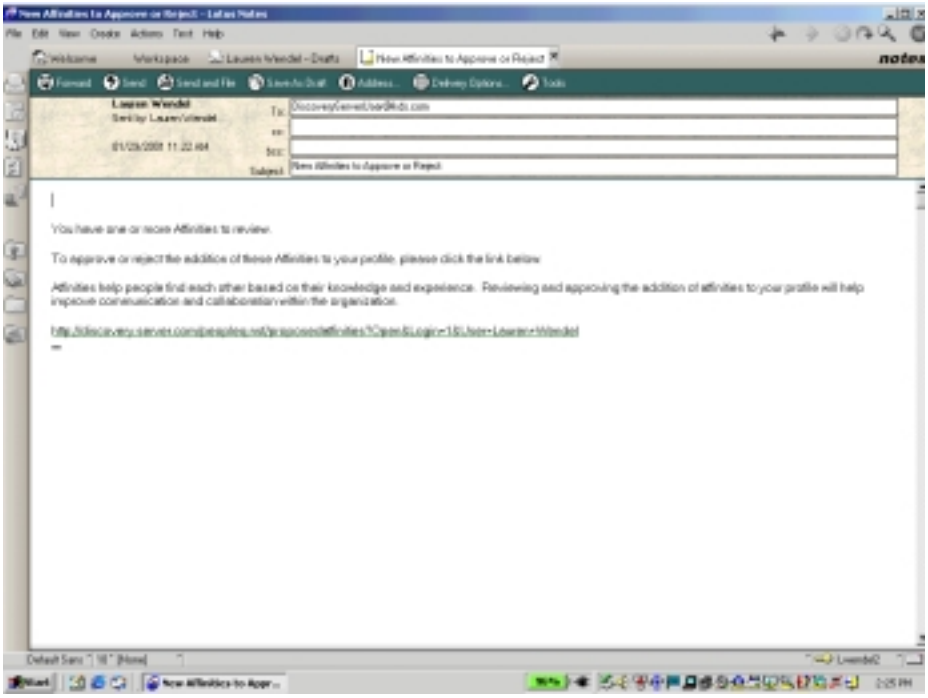

Figure 16: Automatic notification of discovered affinities via e-mail.

To review their own Expertise Profile, end users simply select the correct link. The Discovery Server then automatically presents them with options to Approve or Reject the proposed term for publication. Depending on Discovery Server configuration settings, the end user may see additional options which let them defer this decision for a specified period of time (see Figure 17). Affinities approved for publication to the Expertise Profile are listed in descending order, strongest affinities first.

Figure 17: Approving proposed affinities.

End users can access and revise their profiles at any time. If a user has given the Discovery Server permission to scan their e-mail inbox to assess the relationship of e-mail content to public K-map category areas, he or she can revoke that permission at any time. This is in keeping with one of the Discovery Server's principal design goals: that end users always retain control of information published about them in their Expertise Profile, and made available to others through the K-map.

Knowledge Audits

As touched on above, administrators can generate system-wide reports designed to evaluate K-map content usage patterns, measure the value of organizational data sources and identify areas of expertise within the organization. The Discovery Server enables organizations to monitor trends in these areas over time, via its Metrics Administrative report capabilities. In addition, using the Discovery Server API Toolkit, it is possible to augment metrics information by including additional metrics data to be stored within the Discovery metrics data store or to influence metrics weighting algorithms that are subsequently processed by the Discovery Server Metrics Calculations to provide valuations of information content and subject matter experts.

The Discovery Server provides a set of standard Metrics Administrative reports to monitor usage patterns across the following areas over time, and a report generator to create other customized reports from Discovery Server metrics information:

- Most Active K-map Categories (with associated number of Affinities/Category)
- Most Active/Valuable Documents
- Most Active Document Authors (with associated document values/Author)
- Most Active/Read Documents
- Most Active/Responded to documents
- Most Active/Linked to (cited) documents
- Most Active/Linked from (cited) documents

Metrics Administrative reports provide an accurate report of the value of organizational data, and also can be used to identify areas where knowledge or associated expertise may be deficient (see Figure 18).

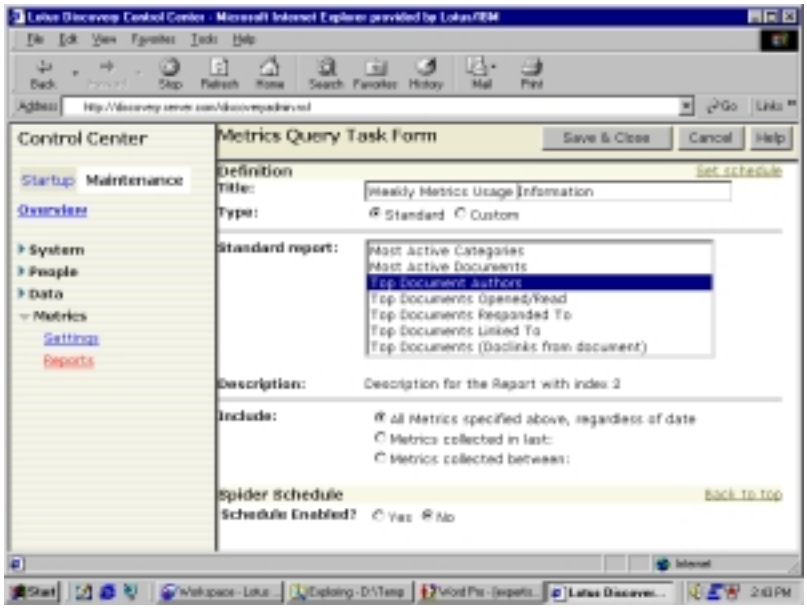

Figure 18: Assessing the value of information with Metrics Administrative reports.

Conclusion

While technology can supplement and support knowledge sharing among individuals, technology alone is insufficient to create a "knowledge culture." To successfully leverage knowledge discovery systems, organizations must actively evangelize and champion the sharing of ideas and knowledge in the workplace. Helping colleagues across the organization should be explicitly rewarded, for example.

As one IT manager put it, "You can't just build one of these systems and expect it will be used just because you build it. The hard work is the process and cultural engineering that needs to take place alongside the technical development. If you've got that kind of culture and systemic process, adding automated tools is easy. But the organization has to be prepared to engage in that kind of collaboration and sharing."

The Discovery Server offers Expertise Profiling capabilities that provide automated, accurate valuations of content, interactions and expertise across the organization. With built-in controls that respect users' privacy, and give them control over the data about them that is made public, the Discovery Server supports those who are new to the knowledge culture, as well as those users who have experienced the benefits of information sharing and relationship building through knowledge discovery solutions.

For More Information

For more information on the Lotus Discovery Server, contact your Lotus or IBM Sales Representative or visit the following URLs:

Knowledge Management	www.lotus.com/km
Lotus Discovery Server	www.lotus.com/discoveryserver
WebSphere Portal Server and K-station	www.lotus.com/kstation, www.ibm.com/wps
The Lotus Notes Network and Iris Today webzine	www.notes.net www.notes.net/doc www.notes.net/kdsforum
Lotus Development Corp. general information	www.lotus.com

The following resources are available for download at www.lotus.com/discoveryserver:

- The Lotus Discovery Server: Taking Advantage of the Collective Experience in Your Organization (white paper)
- Building Enterprise Taxonomies with the Lotus Discovery Server (white paper)
- Lotus Discovery Server product spec sheet
- Knowledge Discovery in Action Business Brochure
- Lotus Discovery Server Reviewers Guide
- Lotus Discovery Server API Toolkit

The following resource is available for download at www.redbooks.ibm.com:

- Inside the Lotus Discovery Server

The following resources are available for download at www.lotus.com/kstation:

- Lotus K-station Portal Overview (white paper)
- Lotus K-station Reviewers Guide
- Lotus K-station product spec sheet
- K-Station Performance Evaluation Guide (white paper)
- K-station Free 90 Day Trial Download