



The Discovery Server Team

Interview by
Tara
Hall

Level: All
Works with: Discovery Server
Updated: 03/01/2001

Before you know if you need knowledge management in your organization, you need to know what it is and what it can do. This month's interview with members of the Discovery Server team explores Lotus's Knowledge Discovery System, focusing particularly on the recently announced back-end, the Discovery Server.

Also for the month of March, members of the Discovery Server team will be answering your questions about the Discovery Server in the [Iris Cafe Developer Spotlight Forum](#) here on Notes.net.

With so many different knowledge management offerings available today, how does Lotus differentiate its knowledge management offering from those of other companies?

Wendi Pohs

Other companies focus on specific areas of a knowledge management solution, like search or categorization or portal creation. Lotus's Knowledge Discovery System (KDS) provides both the infrastructure and the tools needed to build a complete solution—from aggregating information from different sources to giving users the ability to find the information they need and then to collaborate and take action.

Dave Newbold

One of those tools is K-station, a knowledge management portal that is a portal into your application space with many unique features. One of the collaborative features is people awareness. K-station is aware of who is in the portal at the same places that you are in and aware of the people membership of a place.

Another unique feature is re-use of a place. A place represents a group of people and a group of applications. A good example of how a place might be re-used is by a sales team in an organization that needs to respond to a customer on a request for proposal (RFP). This is a task that sales teams do all the time, so it's a reusable kind of object in knowledge management. The whole idea of knowledge management is to provide re-use wherever possible.

Let's say the sales team wants to respond to a plastics RFP. There may be an engineer, a marketing person, a lawyer, a customer liaison, and a manufacturing liaison who have responded to the same RFPs in the past. When the sales team responds to a new RFP, the team can re-use an existing place template. Bingo! The team has the right database repositories, the right people, and usually, the right process model in place to re-use.

K-station is also integrated with the Discovery Server. The Discovery Server provides search, browse, and expertise location services to the K-station portal and other e-Business applications.



Clockwise from 11:00: Dave Newbold, Wendi Pohs, Lauren Wendel, and Jaye Fitzgerald

So how does Discovery Server fit into the Knowledge Discovery System?

Dave Newbold

The Knowledge Discovery System (KDS) is the name for the combined product offering of K-station and the Discovery Server.

Wendi Pohs

The Discovery Server is the back-end of the Knowledge Discovery System. Its services—the spiders, K-map Building service, K-map Indexing service, and Metrics services—access, manage, and analyze the information from a variety of corporate and external sources. The Discovery Server's user interface provides search and browse access to information from these sources.

The K-station portal front-end is more focused on personal, community, or team activities. K-station is used to aggregate the information you know you need to know; the Discovery Server provides access to information you may not know you need.

So just what is the Discovery Server and how does it work?

Dave Newbold

The Discovery Server is a back-end server that spiders documents and your organization's directory to create a catalog, or K-map, of documents, expertise, and places that the end user can browse and search. When used with K-station or QuickPlace, we automatically categorize these virtual workspaces in the K-map.

Another component is what we call metrics, which is a computational program that looks at existing documents and relationships between documents and people. The metrics component does two things. First, it calculates the value of the document, and second, it calculates an affinity between a person and documents, based on their interactions with those documents, that helps produce the expertise affinities.

The Discovery Server has unique features that have never been seen before. One feature is automated tools for assisting with the creation of the catalog. We can generate a catalog automatically, or at least a first draft of the catalog automatically. We spider a representative selection of your documents or databases, generate the catalog, and provide a full-featured, powerful editor

to edit the titles, structure, and configuration of the taxonomy.

Related to that, we also provide an expertise locator, which unlike other "people finder" systems, is a fairly automated system. When you load the Discovery Server, we pull the existing user profiles from the directory, generate a separate profile database, and calculate affinities between people and the documents they used and authored. Then the Discovery Server automatically generates the affinities and proposes them to the end user for approval. That level of automation is not available in any other product. We give you an affinity to what is essentially a rarefied and edited category, not just a set of keywords. That's another big bonus.

We expect most people to use the Discovery Server the way we've shipped it with K-station. However, we have a lot of business partners who want to take the underlying functionality and embed it into their existing applications. For example, integrating it into a Web site to support communities or integrating it into a sales application to make the application smarter about the information that you have on a customer and on the relationship with the customer. One of the biggest holy grails of knowledge management is discerning who knows what.

A favorite quote about knowledge management from the director of research at a major automotive firm is, "If we just knew what we know, this organization would be 30 percent more profitable." That's millions of dollars in profits. If you can do that for a global organization, that's a big win.

There are several terms you use that seem to have a specific meaning in the knowledge management space—spidering, affinities, taxonomy, for instance. Can you give a quick definition of them?

Wendi Pohs

A taxonomy is a hierarchical set of categories and documents used to find information—we also refer to it as the catalog or K-map. Spiders are multi-threaded processes that collect data. The K-map Building, the K-map Indexing, and Metrics services act on the data. The Metrics service also keeps track of what users do with the documents. Affinities are relationships between people and categories in the K-map.

With a growing need in organizations for knowledge management, why hasn't there been a product like the Discovery Server available before now?

Wendi Pohs

I think that's a cultural issue rather than a technical issue. Organizations have only recently begun to recognize that they can become more efficient by providing general access to information created by different functional groups. Hardware is less expensive and software has become more sophisticated; it takes a lot of cycles to make search look this easy. Also, the ubiquity of the Web search engines on the Internet has made knowledge management technology more familiar to people who expect to be able to use it internally at work.

Dave Newbold

A lot of the technology we're using has been around for over 20 years. What's happening now is sort of a confluence of events. As Wendi mentioned, we now have computers that are fast enough, the hard drives that are big enough, and the need that's overwhelming enough to show us that there's a real opportunity here.

I think there aren't many other examples in the marketplace of companies trying to carry automation to this level, and we were drawn to it by studying what our partners and customers had done. Our customers have done some amazing things similar to this, but usually in a very manual process requiring a lot of people and time. We felt that there was an opportunity to automate a good portion of that process. It may not be completely automated in the first

release, but we can perfect that. Some automation is a lot better than no automation.

Does the Discovery Server leverage any Domino capabilities?

Dave Newbold

Yes, Domino provides some rich capabilities that we leverage in the Discovery Server. For example, we synchronize our security and profile generation with the Domino Directory and also use the DIIOP, LDAP, and HTTP services provided by Domino. We expect that in subsequent releases, the Discovery Server will support more environments. We'll always be using Domino underneath the covers, but even today, there's certainly no architectural dependency. And there's no requirement for a Notes client. We use the services of Domino for convenience more than anything else. I should also mention that Domino environments typically have the richest content and the most consistent metadata, like titles, author attribution, etcetera, which makes our job a lot easier.

How does the Discovery Server integrate with an organization's existing systems and products?

Dave Newbold

The Discovery Server can enhance almost any application that uses the Web formats and protocols. We use HTML and DHTML for the client. We use Java as an API for its platform capability. On the back end, we use CORBA, but we don't expose it to applications. We are very standards-based.

Our customers and partners are very excited about integrating the Discovery Server functionality into their applications. We have a Java API for client access and a Spider SPI in development. I think you can integrate just about any product with the Discovery Server.

Already, we have integrated Sametime and QuickPlace support as well as spider support for Domino.Doc. We can integrate closely with Domino Workflow, and we anticipate that we'll integrate with other Domino-enhancement technologies. The Discovery Server can also access information from back-end systems and relational databases through Domino Enterprise Connection Services (DECS), which comes with Domino.

For organizations that want to implement Discovery Server, what type of support team do you recommend? Obviously, organizations will need server administrators, but do you see a need for librarians or taxonomists to help organize information?

Wendi Pohs

Organizations need to do some up-front work to use the Discovery Server to build knowledge management applications. If the applications are narrowly focused, like accessing research and engineering reports, then internal content experts can do the required analysis. But it's probably a good idea to use a librarian or taxonomist as the knowledge management applications grow.

Librarians are skilled in performing information audits, which are the knowledge management equivalents of systems analysis and are used to determine the information needs of organizations. Based on these audits, librarians select appropriate sources to include in the systems and also choose appropriate terminology to label categories so users can easily find the documents they need.

The Discovery Server learns from what the librarians do, so librarians work to train the system, rather than individually indexing large groups of documents.

Dave Newbold

Although we have gone to great lengths to simplify the installation, configuration, and administration of the Discovery Server, customers and partners need a Domino administrator to prepare the environment for KDS.

Once you install KDS, you must decide which resources—databases, Web sites, file systems, and so on—are representative of the information in the organization and what information you want to use to create the initial catalog.

Consult with subject matter experts in your organization to refine the catalog labels and structure. As Wendi will tell you, this is the domain of information science. There are many specialists available to help. One of the options Lotus will provide with a KDS purchase is a Jump Start package that includes the consulting services needed to survey your information environment and create the catalog.

We are providing a lot of support from the product group. Wendi is an ontologist with a background in information science, who's been working on our team for a while. She's writing a book about how to make the best use of the Discovery Server. The book will be available from [IBM Press](#) later this spring. She's published articles and been to a number of trade shows. I recommend that organizations that don't have taxonomists, ontologists, or librarians on staff take a look at these resources.

Web sites like [About.com](#) and [Yahoo](#) use humans to categorize information on the Web and many companies are manually categorizing their internal information. What advantages and disadvantages does the Discovery Server have over systems that use humans to manage information?

Wendi Pohs

The Discovery Server provides a combination of both human and machine-assisted search and categorization. Its K-map Indexing (or full-text search) module is completely automatic and is good for the impatient user who knows his subject well. But the Discovery Server also provides tools to support the creation of categories and automatic categorization of new information.

Using the K-map Builder and the K-map Editor, human indexers can categorize large numbers of documents relatively quickly. The K-map Builder learns from what humans do. No information retrieval system is perfect because the "right" answer depends both on the question asked and on the need of the user asking the question. Totally manual systems, like Yahoo and About.com, are constrained by time and by the biases of the human guides. The Discovery Server allows server administrators and taxonomists to decide how much or how little automation they'll use based on the information needs of their users.

Dave Newbold

Automation is basically picking up the digital bread crumbs, if you will, of what an organization's users have been doing with their documents. The Discovery Server replays all the micro-decisions that an organization and its users make. For instance, do users alter their documents? Do they forward mail documents? Do they create category links to documents? Do they respond to documents in a discussion database? Do they delete documents? All of these actions add up to defining which documents an organization values the most, how much organizations value them, and what affinities users have to that information.

The Discovery Server does the accounting and the filing for you. But you as the human make the value decisions. The Discovery Server can't make value decisions. It can only add information to the catalog. The human element is always there. This is not an artificially intelligent program. We're just using the computer for what it's good for—storage and addition.

Glen Kelley

A major advantage we have is being the first to offer the ability to find "everything about" a given topic including people, places, and documents

categorized into one catalog. Relying on humans to keep profiles up to date has proven to be neither easy nor successful for most organizations, so this is an advantage over relying on traditional methods for categorizing and finding experts.

The Discovery Server analyzes user activity, including e-mail messages. What steps have you taken to ensure users' privacy?

Dave Newbold

The general principle we have about the Discovery Server is that the user is always in control of the representation of himself and his information. The Discovery Server does not publish affinities or make private information known to the world, unless the user gives it explicit permission to do that for every piece of information. We absolutely want to protect the user's privacy.

We should also mention that we allow e-mail to be spidered, but e-mail is only spidered to find affinities when the user doesn't have enough publicly authored documents to generate affinities. We don't publish e-mail messages into the catalog or allow anyone to index or search e-mail. That isn't the point of monitoring user activity. The point is to find out which subjects the user finds valuable. We're providing affinities and again, we give the user the opportunity to approve or deny affinities.

Lauren Wendel

The Discovery Server has been architected to respect the end user's requirement for privacy and control of information. It's essential that end users be notified that information is being evaluated about their interaction with business content. As Dave mentions, profiled users maintain control of affinities or other skills information published to their profile by approving any category terms the Discovery Server determines and proposes that indicate their affinity to public content category areas.

Specifically, end users can edit their profiles, approve or deny affinity terms for publication, and control whether their e-mail is evaluated by the Discovery Server to assess the relationship of e-mail content for affinities to the organizations public categories maintained by the Discovery Server.

How does the Discovery Server use the information it gathers from analyzing data?

Dave Newbold

It uses the raw usage data, or metrics, to create document values and to create affinities. Document values represent the calculated sum of the users response to the document, such as citations, forwarding, response documents, reading, etcetera, and indicate a document's general value to the organization. In the K-map interface, documents are listed by this value metric.

Affinities are calculated by looking for evidence of a person's relationship to a category in the content catalog. If the user has authored a document that is categorized, has read a lot of documents in a category, or has cited many documents in a category, for instance, then the metrics component proposes an affinity to the user. If the user accepts, then the affinity is published into the catalog and to the user's profile. Affinities are dynamic and decay with time. The people with a category affinity are ranked by the strength, that is sum of the evidence, of the affinity.

Profiles contain information about users' projects, skills, job type, and other information, including affinities, which have been mentioned several times. What exactly are affinities and who controls them?

Wendi Pohs

The Discovery Server builds and maintains user profiles in a repository that users can query directly to locate experts by skill, experience, project, education, and job type. The Discovery Server creates profiles in several ways: either by drawing demographic data from any LDAP server or Domino

public directory, or by mapping fields from other, specific applications such as team rooms, discussions, and project tracking. The Discovery Server then uses an affinities mining tool to determine relationships between known categories and user activities.

Lauren Wendel

As mentioned, affinities are calculated by the Discovery Server metrics processes, which collect and calculate data about users interactions to all documents clustered within a K-map category. This affinity relationship is measured based on the observed behavior of the person in the information environment. Actions, such as authoring a document, responding to a document, editing a document, creating a link to a document, are metrics data that the Discovery Server monitors. Aggregate individual activity to documents within content categories is calculated to contribute to the affinity score and to the expertise of the individual.

For example, as I author documents in various support, sales, and technical databases within the K-map content category areas "Discovery Server" and "Expertise," the Discovery Server spiders those repositories, collects my usage interactions, and calculates a strength of affinity based on my contributions and responses to documents in those categories.

When the individual's interaction with Discovery Server K-map content categories reaches a certain threshold (relative to the interactions of all other people tracked), the Discovery Server sends an e-mail notification to the end user with that proposed affinity information. The e-mail notifies the end user of the affinity determined (proposed) by the Discovery Server and requests confirmation and approval or disapproval to publish the identified affinity. If approved, the affinity appears in the user's profile document, accessible to others through K-map search. In that way, the end user has complete control over what information indicating their expertise is published to their profile and made available to others searching for subject expertise.

The Discovery Server undergoes a security review by the IBM security team at the Watson Research Center, known as the Ethical Hackers. How does the Discovery Server handle security, and what steps have you taken to secure information?

Dave Newbold

Essentially the IBM security team reviews the implementation for security flaws and provides the product team with advice. At our review level, they're looking at the architecture and the security mechanisms that we have in place to ensure security.

The concerns we have with security and security information are numerous. The Discovery Server needs to respect the security of the original source document, meaning to make the server work, you actually have to give the Discovery Server fairly global rights to your data. To do that, you have to become sort of a super-user on that end system. We take that responsibility very seriously, because we're going to be copying that data, processing it, and displaying portions of it as a result. We want to make sure that only those people who have access to the original data will see any of those results and to maintain the integrity of the source data in the process.

To that end, we have controls for encrypting work queues and making sure that the server-to-server communication is secure. In addition, we don't want any replay attacks from unauthorized users going into the server.

Even today there are potential vulnerabilities based on the possible configurations, which we discuss in the documentation. Before organizations implement Discovery Server, we recommend that administrators carefully review the documentation to understand some of the assumptions we make about the environment. For instance, we assume that the servers are in a physically secure location and that the interaction between servers is also secure. Having said that, security is always a fairly complicated issue with a

lot of different permutations.

What advice do you have for administrators concerned with performance and capacity planning?

Dave Newbold

The first advice is to read the documentation. We're currently documenting the performance characteristics that we've seen with the product. A performance and capacity planning guide will be available after shipment. We're actually in the throes of getting the product wrapped up, so we don't have the final performance numbers yet.

For those who want to customize their environments, what options are available to them?

Jaye Fitzgerald

The end user has several options for customizing the K-map user interface (UI). They can customize the font size used in the K-map UI. The font choices are similar to those available in the browser. End users can customize presentation of table data within the K-map UI. The Document, People, and Places tables can be changed to display different fields from the default fields. Finally for administrators, we will be publishing the steps necessary to update the K-map UI with their own images.

Dave Newbold

The steps that Jaye mentioned describe how to modify the K-map UI. The HTML source is in the install kit, and you can extend it by adding your own branding and changing the background bitmaps, and so on.

In addition, we will have a Java API available to extend and customize the Discovery Server by using it as a back-end service and giving it whatever look-and-feel you like. We have some business partners, like [The Brain](#), who are already working with different visualizations, and those visualizations are a very interesting and useful customization of the product.

Willie Arbuckle

The API is known as the KDS API Toolkit. It contains classes for customizing the search, metrics, taxonomy, configuration, and administration services. For instance, you can create queries against data collected by the KDS, automatically register content from outside applications for spidering, and create your own metrics to influence expertise location. In the first release, the API will support Java applications, servlets, and applets.

Looking ahead, what role will the Discovery Server play in Lotus's knowledge management strategy?

Glen Kelley

The Discovery Server is a key component of Lotus's knowledge management strategy. It is the most comprehensive knowledge management enabling technology ever introduced. We're helping individuals and teams to find comprehensive insights and answers and to access experts in their organizations to solve everyday business problems. We see these capabilities enriching existing and future applications, while forming the basis for many different business partner solutions moving forward.

Dave Newbold

The Discovery Server has a pivotal role because it provides back-end services that can bring together a lot of the technologies that we've talked about over the years. Those technologies include knowledge portals, like K-station, in which the Discovery Server is the innovation engine for finding and providing context to the communities of practice that have been exposed in the portal.

The portal certainly stands alone, as do all the other Lotus knowledge management products like Domino Extended Search, Domino.Doc, and Domino Workflow. They have their own value propositions, but I think they

work a lot better when they're integrated. The Discovery Server is one integrating technology for bringing those products and their resulting data together.

I think the Discovery Server has a great future. We've had a lot of great customer response, both in terms of interest in buying as well as in the beta program. This product has the potential for accelerating the evolution of e-business.

During the month of March, members of the Discovery Server team will be answering your questions about the Discovery Server in the [Iris Cafe Developer Spotlight Forum](#) here on Notes.net.

ABOUT WILLIE ARBUCKLE

Willie Arbuckle is an architect in the DevTools Group at Lotus, and has recently begun working on defining and implementing the API for KDS (Knowledge Discovery System). He also has responsibility for the LSX Toolkit. Willie has been at Lotus for 14 years. Prior to developing the LSX Toolkit, he worked on the Notes back-end classes for LotusScript. He has also worked on the Version Manager in 123/W and on the 123/G for OS/2 team. Before joining Lotus, Willie taught computer science courses at the University of Ulster, in Northern Ireland. Both his bachelor's and master's degrees from the University of Ulster are in computer science. When not writing code, Willie teaches Irish Gaeilge classes in the Boston area.

ABOUT JAYE FITZGERALD

Jaye Fitzgerald is a principal software engineer and lead developer for the K-map user interface. Prior to joining the KDS team, he contributed to R5 as a member of the Notes client team. Before joining Iris, Jaye worked as a software engineer at Wall Data for two years and OneSource Information Services for over ten. Jaye has a BA in management from Boston College and a graduate certificate in computer science from Harvard University. Jaye spends his free time with his wife and three sons. He is an active Scout leader and sport coach and enjoys fishing, camping, skiing, and playing basketball.

ABOUT GLEN KELLEY

Glen Kelley is group manager for Discovery product marketing. With more than a dozen years of high-tech marketing experience, Glen is now involved in positioning the Lotus Discovery Server as the market leading enabler for expertise location and unstructured data technology. In his role of defining and assessing the knowledge management marketplace, Glen is educating today's organizations on the benefits of advancing technology that enhances the relationship between people and the content they interact with. In addition, Glen is rolling out the Discovery market strategy along with expanding the market with a position of leadership. Prior to joining Lotus, Glen was a principal with marchFirst (formerly Whittman-Hart), focusing on solutions in the areas of collaborative applications, messaging and CRM solutions. Glen works at Lotus's corporate headquarters in Cambridge, MA while residing in Minneapolis, MN. He has a bachelor's degree in music business from Western Illinois University.

ABOUT DAVE NEWBOLD

Dave Newbold is the Iris general manager responsible for the Lotus Discovery Server. Dave came to Iris in 1993 to integrate Internet protocols into Notes, resulting in Web Navigator and the InterNotes News gateway. After a start-up stint organizing the Notes.net site, he brought Domain Search to life and started work designing the Discovery Server. Prior to Iris, Dave worked on networking products for Lotus, 3Com, and NYNEX.

ABOUT WENDI POHS

Wendi Pohs is a principal taxonomy specialist at Iris and the author of an upcoming book on KM methodologies. Wendi joined Lotus Development Corporation in 1996 and has worked on various projects as a spec writer, online help designer, and user assistance manager. Prior to joining Lotus, Wendi worked at the American Mathematical Society and at Digital Equipment Corporation. Wendi received her BA and MILS degrees from the University of Michigan.

ABOUT LAUREN WENDEL

Lauren Wendel is product manager for KDS - Discovery Server and expertise technologies. Prior to joining Iris, Lauren worked with the Lotus Enterprise Integration

team for five years, overseeing the initial releases of Lotus Enterprise Integrator, DECS, ERP Connectors, and the Connector API Toolkit. She has also worked as a developer consultant within the Lotus Business Partner program, and previously within the 1-2-3 engineering team. Lauren's also managed systems planning at Wells Fargo Bank, Citibank, Duke University School for Executive Education, and Grant Thornton Ltd. She enjoys running the "occasional" marathon and sings with a community chorus.

ABOUT TARA HALL

Tara Hall is a senior user assistance writer for Lotus where she has worked for over two years. She is part of the Web Applications team and writes online help, programming guides, and release notes. She also is a member of the Notes UA Web team.