

**Level:** Intermediate  
**Works with:** Discovery Server  
**Updated:** 04-Jun-2002

## Creating meaningful K-map taxonomies

by  
[Dick McCarrick](#)

Among knowledge management professionals, a fine-tuned, easy to understand taxonomy is just about the most beautiful thing in the world. After all, here is a visual representation of that most elusive yet critical concept, *what your organization actually knows*. With a well-constructed, meaningful taxonomy, users can quickly find the right people and information they need to achieve their goals. It's equivalent to having instant expertise at your fingertips, ready to answer questions you may have at any time. Sounds wonderful, doesn't it?

Of course, in the real world, it's historically been very difficult to construct a useful taxonomy of corporate information, especially in larger organizations. Indeed, before the advent of computers, it was all but impossible, particularly for major companies where employees came and went and new technical documents were constantly added to the corporate store of knowledge. So in most situations, taxonomies tended to either be superficial and general, or constantly out of date—often both.

Lotus Discovery Server helps change all that. Now you have a state-of-the-art tool that can keep track of your documents as well as your user's skills, regularly checking for updates and additions. More importantly, Discovery Server uses complex mathematical metrics to analyze the content of your documents and the experience of your employees, find affinities between them, and place them into categories. These categories comprise a computer-generated taxonomy, or in terms of Discovery Server vocabulary, a *K-map*.

But no computer program, however sophisticated, can precisely predict exactly how a particular organization may want to structure its content. So to build a truly meaningful, easy to use K-map taxonomy, humans and computers must work together. This requires some time, patience, and creative thinking (on the part of the humans anyway). But if done correctly, the results are more than worth it.

This article offers advice on building useful, robust K-map taxonomies. It presents things to consider when deciding which type of taxonomy scheme works best for your organization and who should be involved in the process. We'll prepare you for all the major steps involved in planning and creating a taxonomy. These include:

- Forming the K-map taxonomy team
- Selecting your taxonomy schemes
- Setting up Discovery Server to create a preliminary K-map
- Running Discovery Server K-map Building Service (also known as K-map Builder) to create a preliminary K-map
- Reviewing the preliminary K-map and changing Discovery Server setup parameters to create a revised K-map to edit
- Editing the K-map taxonomy with the K-map Editor

Please note that this article is not a step-by-step tour through the Discovery Server interface. Instead, it's intended as a best practices collection of tips, techniques, and customer experiences. [Click here](#) to view Lotus Discovery Server product information. For [Discovery Server documentation](#), see the Lotus [Documentation Library](#). And

for more information about Discovery Server and knowledge management in general, consult *Practical Knowledge Management: The Lotus Knowledge Discovery System* by Wendi Pohs, which is available from [IBM Press](#).

## Forming the K-map taxonomy team

One of the more distinguishing characteristics of the human intellect is our ability to join ideas—some view this as the very definition of creativity. This is what a taxonomy is, a representation of ideas and how they are joined, how they relate to each other.

As much as we in the industry are loathe to admit it, even the most powerful computer software can't really join ideas the way people can. In fact, it can't even truly grasp the concept of what an *idea* is. It can't read a document and form a vision of what it's really about, what its author intended to convey to readers. To computer software, a document is little more than a "bag of words."

But what computer software *can* do, it can do very quickly. For example, Discovery Server can examine thousands of documents far faster than a team of people can. It can record each word in a document, how often it appears, and whether certain words tend to appear near each other in a statistically significant way. And it can create categories based on these words, build a hierarchy from these categories, and place documents with similar content into each category. This is no mean feat—just consider doing it "by hand." Imagine reading each of your organization's documents in all their forms, keeping a running tally of how often certain words appear in each, applying highly complicated mathematical formulas to the results, and then classifying documents accordingly. Most people would find this work long, hard, and tedious. Fortunately, Discover Server does all this for you.

But remember, even when armed with Discover Server, a computer can't think. The categories it creates may make perfect sense in an abstract world ruled by numbers. But they may not be instantly self-evident to users who have to navigate through them. So in all likelihood, you'll want to configure Discovery Server in a way that prepares it to produce the kind of K-map most useful to your site. And you'll need to edit the categories on this first K-map. To do this, you'll need the right experience and expertise on hand. And if yours is a site with a large and complex body of corporate knowledge, you should consider forming a K-map taxonomy team

### Important K-map taxonomy team skills

In general, your team should possess both technical knowledge of your content and experience in creating meaningful taxonomies. For example, recently a site analyzed its resource needs for planning and creating a K-map of its stored documents. This site possessed a particularly broad and complicated information trove, encompassing numerous subject areas across multiple departments. As a result, the site concluded its taxonomy team needed the following roles:

- Chief content specialist to oversee all cross-functional content and decide which content to include in the K-map
- Individual content specialists/knowledge stewards familiar with the content of each area within the site
- Taxonomy editor familiar with the content and how to organize it
- Knowledge management architect/taxonomist to guide the organization/classification/taxonomy building process

Don't panic when you see the above list—this doesn't represent all the people you now have to hire to build a usable K-map. It merely suggests roles in the taxonomy planning and building process, not permanent job descriptions. In most cases, you'll already have the people on your staff to serve these roles. Also, this list represents only one organization's requirements for a taxonomy team. Your site may choose to populate its team differently.

### The role of the taxonomist

One common thread among successful teams, however, will be at least one *taxonomist*, a person who knows about constructing taxonomies. "Taxonomist" is a profession that probably mystifies most of us. So in identifying the right person to assume the role of defining a common taxonomy architecture that can be used at a departmental and enterprise level, you might want to consider the following qualifications:

- Experience designing effective classification schemes that address all metadata requirements for the creation, storage, management, and retrieval of intellectual assets
- Experience designing overall content management strategy, including defining the roles and responsibilities of all vested parties (author, indexer, end user) around knowledge management work flow
- Ability to define the change management processes behind the taxonomy to manage content over time
- Skill in people, process, and technology aspects of information architecture, data administration, and information source management
- A Masters Degree in Library Science, which is recommended

In addition, you'll need at least one person who knows about and understands the technical content of your knowledge. These two human skills—taxonomist and subject matter expert—working closely with each other and in conjunction with Discovery Server, will be crucial to the process of organizing and making sense of all your raw information.

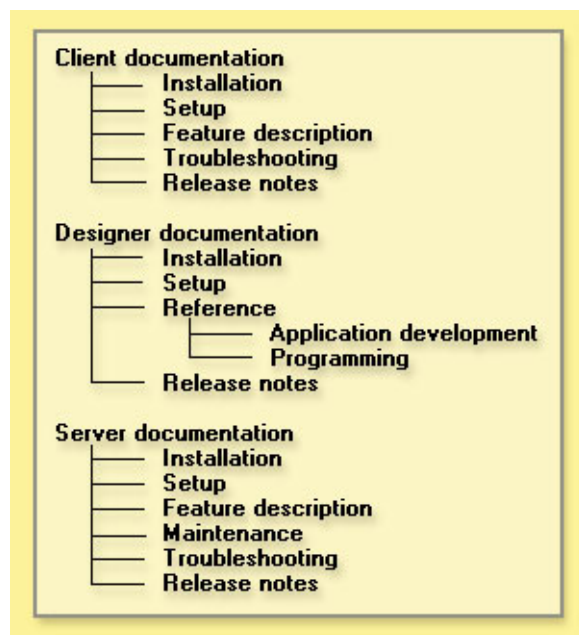
## Selecting a taxonomy scheme

A lot of people assume that for any given collection of items, there's one best way to organize them—in other words, a place for everything and everything in its place. (Perhaps this is residue from that biology class we took in high school.) But reality is usually more complicated.

### Types of taxonomy schemes: rules-based versus natural

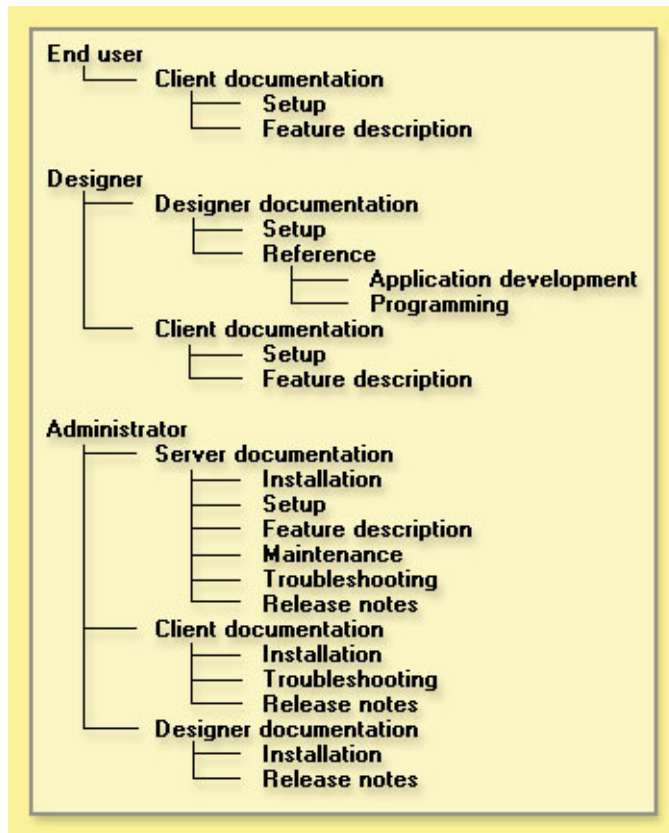
For example, consider the task of creating a taxonomy of documentation for a complex client-server software system. This unnamed product consists of client, designer, and server components, and the documentation includes on-line help and printed materials. It covers installation, setup, feature descriptions, maintenance and troubleshooting, and release notes for all components.

The initial impulse would probably be to organize the taxonomy along product lines, as shown below:

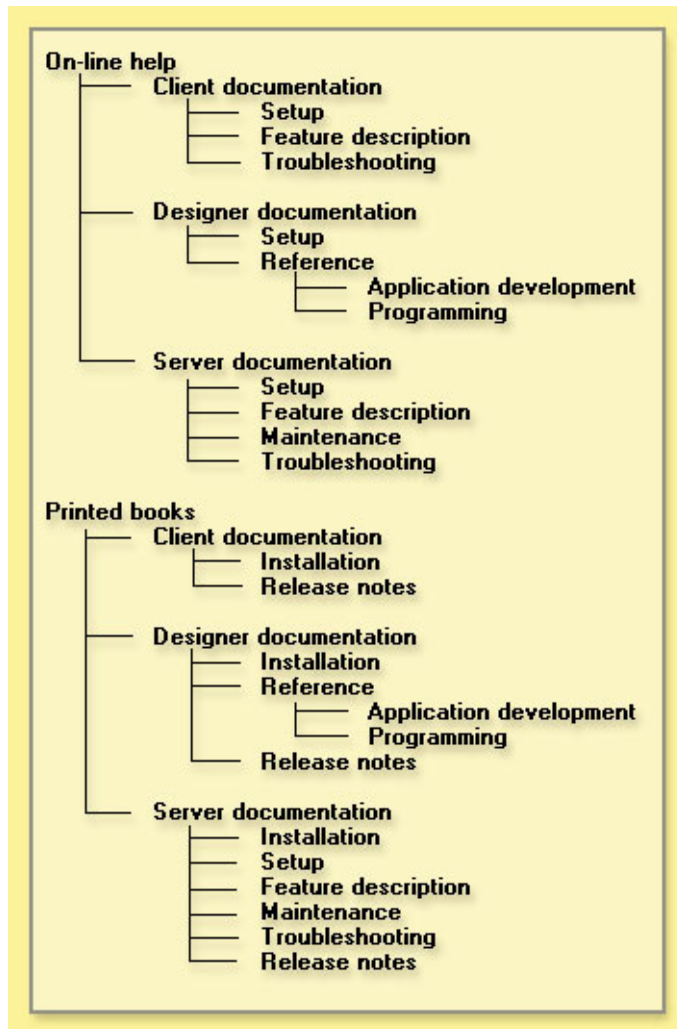


This is an example of a *rules-based* taxonomy. (Another, pre-computer example is the classification system used by many libraries.) This is a traditional approach, based on predefined rules and therefore, structured. This structure makes rules-based taxonomies easy to understand and describe. But in practice, they're not always easy to implement, or keep up-to-date, especially when your organization adds content that doesn't readily fit into any one category (or even worse, seems to apply to them all).

Although rules-based is a perfectly legitimate method for organizing your documents, it's not the only way. Nor is it even necessarily the best way, especially if you want users to follow this taxonomy to find the information they need, in a way that feels natural to them. In our example, consider for a moment how the documentation will be used and by whom. Some organizations might not want end users to install or troubleshoot their own software, so they assign this task to administrators. Nor do they want anyone other than administrators to read the release notes, since the misapplication of this information might lead to disaster. Also, designers may need to frequently reference the client documentation to better understand end-user needs and to test their applications. So these organizations may prefer instead to structure the documentation taxonomy with a more task-oriented, "natural" approach:



The above example uses what taxonomists refer to as a *natural* classification structure. Natural classification schemes also impose order on a body of information, but they tend to be more informal. Note also that some items appear in more than one category. So is *this* the best way to organize this information? Perhaps, but perhaps not. For instance, you may want users to see at a glance what's available on-line, and/or what's available as printed books (this might be considered a hybrid between a rules-based and a natural classification scheme):



Our simple example illustrates that in the majority of cases, there's no single right way to structure your knowledge. Any number of taxonomies are possible. The taxonomy selected by one organization may not be the best for another. Indeed, a single organization might choose multiple taxonomies for the same data, offering its users several choices for locating information. Consider which taxonomy (or taxonomies) works best for your organization when you plan your K-map. Bear in mind:

- What is the objective of the K-map? The most obvious is to create a way to navigate to the information you need. But K-maps can also offer a graphical representation of what your organization knows, identify "knowledge gaps" in your corporate content, discover redundancy and/or inefficient use of disk space (for example, multiple documents offering essentially the same content), and so on.
- What are the needs of content consumers? Consider how your employees will use these documents, how they think about them and apply them to their jobs.
- What are the needs of content creators?
- Is there an existing taxonomy? You may already have a file/folder structure in place you'd like to preserve and use as the basis of your K-map.
- How clean is the metadata? (*Metadata* is information about your content, such as document title, author, content type, and so on.)
- Is the content suited to automatic classification techniques? Some document management systems (for example Notes) readily lend themselves to automation. Others (for instance Web sites with little text and lots of links) do not.

### Performing a knowledge audit

To help find answers to these questions, it may be useful to perform a *knowledge audit*. Many knowledge management professionals consider this an important step in the taxonomy planning process. A knowledge audit involves identifying potential users of your K-map taxonomy and asking them a series of questions. You should

also record basic format information about your content, including:

- Document size
- Content type (structured or unstructured)
- Internal or external
- Simple or complex
- Single or multi-source
- Existing keyword fields

The answers to these questions will provide the information you need to select your taxonomy scheme. Whichever scheme (or schemes) you choose for your initial K-map, this isn't necessarily your final opportunity. You can later modify your scheme in response to user experience, usability, or additional content. You can also create one or more other taxonomy schemes to replace or supplement your original.

To sum up, K-map planning is an ongoing, creative process that synthesizes the taxonomist's knowledge of how to build a taxonomy, a technical familiarity with your content, and an understanding of human nature and how your people go about doing their jobs. When you complete this phase, you're ready to get started preparing Discovery Server to build your first draft K-map.

## Setting up Discovery Server

Now that you've formed your K-map team and decided upon a taxonomy scheme, it's time to put the computer to work. The first step is configuring Discovery Server's K-map Building Service (also known as K-map Builder, a term we'll use for the remainder of this article). This is an important process, because the better you set up K-map Builder, the less work you'll have to do editing the resulting K-map and working it into usable shape. For information on how to set up the K-map Builder in Discovery Server, see the online Help, which you can download from the [Documentation Library](#).

### Deciding which content to include

One of the first questions you should consider is how much content to include in your first K-map. There are two schools of thought on this. You might select all your content and let K-map Builder have at it. Or you can start with a small subset of your content. The "everything and the kitchen sink" approach takes longer to roll out the K-map. But when you're done, you'll have a complete taxonomy of your organization's knowledge. The "subset" method lets you finish a draft K-map more quickly. But you'll have to keep adding to it until all your content is included.

For most large sites, we recommend starting small. In fact, you may want to include only a relatively limited collection of documents and create the preliminary K-map as a test case rather than a working taxonomy. This allows you to "get your feet wet" with K-map editing and lets you more quickly achieve the encouraging goal of producing a K-map at least some of your employees can use. And as you become more accustomed to K-map editing processes and practices, you'll be better able to complete the remaining work more efficiently.

Another key element is deciding *which* content to include in the initial K-map. If you employ the "start small" philosophy, you might want to include documents specific to one area or department of your organization. This lets you involve users in that department as a pilot site. This can help provide quick feedback on how people use the K-map and how to make it more responsive to their needs. Also, this department can later serve as a resource for others, as you develop and deploy the complete K-map across your organization.

When selecting which content to include in your first K-map, think about:

- What are the current sources of this content? Who owns them?
- What sources do users access most frequently? How do they access these sources?
- What is the life cycle of the content?
- Who can identify the most up to date content?

Also, your organization's content may already be organized in a simple, operating system level folder/file structure. You can import this taxonomy and use it as the foundation for your K-map. For more information on deciding which content to include in your K-map, see the *LDD Today* article, [Selecting and configuring Discovery Server data repositories](#).

### Important K-map Builder settings

Three settings can be very useful when setting up the Discovery Server K-map Builder to create your initial K-map:

- *...set the document maximum per category* at limits the number of documents the K-map Builder can place into a single category. If the number of documents in a category exceeds this number, the K-map Builder creates new sub-categories based on the documents in the original category. This setting is particularly valuable if your preliminary K-map includes categories too large to be easily navigated.



- *Stoplists* contain words you want K-map Builder to ignore when creating categories. Discovery Server comes with a default stoplist consisting of common English words ( *a*, *the*, and so on). You can add words to the stoplist to prevent them showing up in category titles—for example, your company name.
- *Keywords*, in contrast to stoplists, represent terms to which you want to give additional emphasis. The K-map Builder then gives three times the weight to these words. Phrased another way, when the K-map Builder encounters a keyword, it multiplies the number of times it appears in the document by three when creating categories.

We'll talk more about these settings in the next section of this article.

## Creating the preliminary K-map

After you set up Discovery Server, you're ready to run K-map Builder to generate your preliminary K-map.

### Building your first draft K-map

As we noted earlier in this article, the K-map Builder is the service Discovery Server uses to automatically create and maintain the K-map. The K-map Builder builds document categories, creates labels for these categories, and places documents into them. It also identifies documents that don't fit into any existing category. The K-map Builder chooses the three most frequently used nouns (or noun phrases) to label the categories.

When doing this, it applies a mathematical process called vector analysis to the words in your documents. In technical terms, the K-map Builder treats words and phrases in documents as points in a large, multidimensional space. Each dimension corresponds to a single word or phrase and the number of times it appears. When two documents share many of the same words and phrases, they will be relatively close together in this space and will be placed into the same category.

It takes an impressive amount of work to do all this. But once Discovery Server produces the initial draft K-map, it'll require the skills of your K-map taxonomy team to shape the K-map into a final, easily usable form. Of particular importance at this stage will be technical knowledge of the material, in the person of someone who can quickly look at the categories and the content within them and judge whether they possess a true, real-world affinity to each other.

When you first cast eyes on the draft K-map that Discovery Server builds for you, it'll look a little strange. Actually it may look *very* strange, with category names that may range from self-evident to utterly incomprehensible. Some categories may be enormous, with nearly all your documents placed into them. Others may seem to have the same words showing up repeatedly in their names, with little apparent reason. Don't despair; this is perfectly normal.

### Rough tuning with Discovery Server settings

At this point you may be dying to dive in and start editing your K-map. But before you do, you may want to do a little "rough tuning" first by playing with Discovery Server settings and re-running K-map Builder. Then when you've got the draft K-map in reasonably meaningful shape, you can start editing to fine-tune your K-map.

For example, examine the category names for words that don't have any real reason for being there. Recently a customer created a draft K-map and discovered the category names contained long, unidentifiable numbers. They remained a mystery until someone on the K-map taxonomy team recognized that these numbers were codes for various trade publications. As it happened, many of the documents being processed were journal articles. Each had a field for identifying the journal in which it had been published. The K-map Builder discovered these numbers occurring repeatedly, so it obligingly created categories out of them. However, these categories were of limited use in helping users determine what the articles were really about. So we suggested placing these publication numbers on the stoplist so that the K-map Builder would ignore them in the future.

If your K-map consists of a few categories with an exceedingly large number of documents in them, try setting the "...set the document maximum per category at" a lower number. This forces the K-map Builder to attempt to create more categories from this group of documents. You can also use more keywords to help K-map Builder create additional categories, but the overall impact of doing this is relatively small. (Another way to deal with overly large categories is to use the subdivide feature, described in the next section of this article.)

Also look at the "uncategorized" documents, especially if it appears you have an overabundance. The K-map Builder cannot classify small documents (ones that contain fewer than 30 unique "tokens"). So if there's a lot of small documents cluttering up your K-map, remove them.

Feel free to change Discovery Server settings and rebuild the draft K-map as often as you feel necessary, until

you're satisfied you have a reasonably meaningful taxonomy structure. Don't be too concerned yet about category names, even if they strike you as bizarre. You can change them later when you edit the K-map. The important goal here is to ensure you have a decent number of categories to work with, and that in the opinion of your content experts, each category truly represents a meaningful collection of similar subject matter, not merely some artificial mathematical construct. When your K-map achieves that goal, you're ready to start editing.

## Editing the K-map taxonomy

Now you can begin the fun part, editing the K-map taxonomy into a workable, understandable structure your employees can use to find information. You'll do this through Discovery Server's K-map Editor, working closely with your content experts and your taxonomists. The latter will play a particularly crucial role at this stage.

The first thing you might want to do is examine each category Discovery Server has created and determine:

- Is this category real (in other words, is the content therein truly related in a meaningful way)?
- If this category is real, what is the best name for it?

### Determining category validity

The K-map editor helps you answer the first question by displaying the fit values of documents within each category. *Fit values* signify how closely documents in a category resemble the *centroid* document in that category. (Despite its scary-sounding name, the centroid is nothing more than the mythical "average" document in the category. The centroid doesn't actually exist; it's simply a composite model Discovery Server constructs to compare with other documents in this category. You can think of the centroid as the computer-generated equivalent of a Platonic ideal.) The highest possible fit value is 100; a fit value of 50 or more is considered good. K-map Builder assigns fit values when it categorizes documents and builds the K-map. If fit values for documents within a category are high, this implies they are significantly similar and thus the category is likely to be a meaningful one. If fit values are low, you may want to move documents out of this category, or remove the category entirely.

Even if fit values are high, your content experts on the K-map taxonomy team should examine the list of documents within each category to determine whether they appear related. Discovery Server may create a category with high fit values, but base it on meaningless similarities. (Remember our example of journal publication code numbers in the previous section?) If the content experts decide a particular category isn't real, remove it. But if the category does appear meaningful, you're ready to start naming it.

If your K-map categories are large and/or would be more useful broken down into subcategories, you can use Discovery Server's subdivide feature. *Subdivide* lets you instruct the K-map Builder to try to make subcategories from this category. This is a very important tool in the early stages of K-map editing, when many of your categories may be perfectly meaningful but too large and unwieldy to be usable.

Conversely, you may find your K-map contains too many categories to organize easily. In this case, look for categories that may be closely related and merge them.

Another useful technique involves moving documents manually from one category into another. This is especially valuable because Discovery Server can be "retrained" this way. When you place enough documents into a category, the fit values for these documents are recalculated. You can also highlight a category you've edited and request a retrain from the K-map Editor menu. When the retrain process finishes, you'll see the new fit values on all the documents in this category (and its subcategories). Then, when K-map Builder processes new documents with similar content in the future, it will place these documents into this category.

Determining the validity of categories may well be the most difficult part of the K-map creation process. You'll be challenged to find significance in some categories, which may at first appear to be totally random collections. But with a little diligence and creativity (and the right mix of skills on your K-map taxonomy team), you may be surprised how quickly you can start making sense out of these categories. At the same time, don't be *too* imaginative. You may discover categories that really don't have any practical use and should be removed. Don't knock yourself out trying to find meaning in each and every category, insisting there must be some unseen connection just waiting to be discovered with enough mental effort. Finding connections between similar items is creativity, but finding connections where none exists is obsessive and confusing. You don't want your K-map to be either, so don't be afraid to remove categories when appropriate and instruct the K-map Builder to try again.

### Naming categories

Category naming is where art meets science. The category name must represent the content it contains, while at the same time it must be easily understood by users. These names are critical navigational components of the K-map, so you should pay careful attention when assigning them.



For instance, consider the following example, based on a recent customer experience. A large chemical company creates a K-map. One of the category names produced by Discovery Server is "rats/university/drugs." If you're normal, that probably doesn't mean a whole lot to you out of context. But the fit values for this category are relatively high, showing a strong affinity between these documents. After some thought, one of the content experts realizes these are articles that describe new pharmaceuticals that have undergone laboratory testing at a local university. The testing included experimental trials upon (you guessed it) white rats. So this category is renamed "Laboratory tested drugs" and becomes a useful sign post to users looking for this type of information.

As you examine each category, try to work on the easy ones first. At least some of your categories should be obviously meaningful and relatively simple to name, so do these first. If a category looks confusing and requires a lot of thought, you may want to skip it for the time being and save it for later. This lets you make some quick progress in producing at least a partial taxonomy. This gives you some good experience. And it's always encouraging to get something actually done.

Naming categories is an important phase in the development of your K-map because Discovery Server retains each category's name you assign. Then when it encounters new documents with similar content, it places them into this category.

### **Defining connections between categories**

After you have defined a number of categories, you can use the K-map Editor to modify the connections between these categories, and in doing so, define a working K-map taxonomy. This is where you apply the taxonomy scheme your team decided upon earlier. As we mentioned above, a rules-based taxonomy is easier to implement, and for testing purposes this might be a good scheme to follow. But natural classification schemes are generally easier for people to use when finding information. So if you're working on a K-map you plan to widely deploy in your organization, you should consider using a natural scheme.

As you do this, keep in mind how your people will use this K-map. Put yourself in the users' place. Under which categories will they expect to find certain documents? What categories do they consider related to each other? What kind of organization scheme makes sense to them? Navigating this K-map to find information should feel natural and logical to users. Information should be where they expect it to be.

Of course, what seems logical to one person may be confusing to another. It's unlikely your K-map will exactly match how each member of your organization views the world. But you can build in some flexibility. For example, you can create a K-map taxonomy that allows different users to take different paths to the same information. Recall our sample documentation taxonomy described earlier in this article. In our second illustration, we showed how "Client documentation" could appear under both the End User and Designer categories. To do this, you can use *category links*. These allow you to create a category once and then link to it. You can then place these links throughout your K-map taxonomy, where users with differing needs and perspectives can find them. So instead of debating the one perfect place within the taxonomy a category should go, you can, in effect, add it to several places. You can also add the same document to multiple categories to accomplish this goal. However, we don't recommend doing this until your initial K-map is in good working order.

Finally, don't expect to complete your K-map in a single session. The larger the body of content your organization has accumulated, the more involved the creation of your K-map will be. You might build your initial K-map multiple times, tweaking settings and editing the results, until you have a usable, deployable road map your employees can follow to locate the information they need. And even after deployment, you should plan on regularly updating your K-map, as you learn more about how people are using it—and as your organization adds to its collection of content and expertise.

### **Putting it all together**

As you've probably gathered from reading this article, knowledge management involves collaboration—between humans and computers, between different disciplines such as taxonomists and content experts, between formal rules and more "natural" flexibility. All play important roles in that most difficult but rewarding of tasks: making sense of your corporate body of information. In many respects, an organization is defined by what it knows; at a minimum, this represents one of its most treasured assets. Managing knowledge makes for higher-performing teams. And a meaningful, easy-to-use K-map that helps users navigate through this knowledge is a critical component of this process.