

Taxonomies to Tags: From Trees to Piles of Leaves

INSIDE

TAXONOMIES TO TAGS	1
<i>Webogeny recapitulates ontogeny</i>	
<i>The long tail of tags</i>	
The Stand of Trees	6
<i>Dewey Decimal Classification system</i>	
<i>Yahoo!: The inner tree</i>	
<i>Corbis: Controlled words and pictures</i>	
<i>ClearForest grows branches</i>	
From Trees to Leaves	12
<i>OSAF's Chandler: Facets at work</i>	
<i>Endeca: Facing the facets</i>	
<i>Siderean Software: Facets of trees</i>	
<i>PMSI: The market cometh</i>	
<i>Box: NYT.com: All the facets fit to print</i>	
From Trees to Tags	19
<i>del.icio.us: T.as.ty tags</i>	
<i>Flickr: Photo tagging</i>	
<i>Wikipedia: The shape of grass</i>	
<i>frassle: Same tag, different name</i>	
<i>Technorati: Searching for tags</i>	
The Tagging Future	29
Resources & Contact Information	33
Calendar of High-Tech Events	34

Add to the PC Forum folksonomy:
Register for PC Forum 2005,
March 20-22, in Scottsdale, AZ:
www.pcforum2005.com

BY DAVID WEINBERGER

We're delighted to welcome David Weinberger back to our pages – and next month, to a roundtable at PC Forum. A Renaissance man after our own heart, he wrote for us last January about the Semantic Earth; this month, he takes on the trees and forests of metadata. At PC Forum, he will connect a few more dots and co-chair an afternoon roundtable where the assembled multitudes will become a face-to-face festival of attention, metadata, ontologies and real-time emergence of ideas through interaction – with David's crisp mind getting us all to reflect on the patterns we are living and creating even as we discuss them.

The narrative that tells of the first man and woman encountering the tree of knowledge focuses on its tempting fruit. But after we took the bite, we apparently looked up and got the idea that knowledge is shaped like the tree's branching structure: Big concepts contain smaller ones that contain smaller ones yet. Over the millennia, we have fashioned the structures of knowledge in just such tree-like ways, from the departmental organization of universities (liberal arts contains history and history contains ancient Chinese history) to the hierarchy of species. The idea that knowledge is shaped like a tree is perhaps our oldest knowledge about knowledge.

Now autumn has come to the forest of knowledge, thanks to the digital revolution. The leaves are falling and the trees are looking bare. We are discovering that traditional knowledge hierarchies that have served us so well are unnecessarily restricted when it comes to organizing information in the digital world. The principles of organization themselves are changing now that they are being freed from the constraints of the physical world. For example:

{ continued on page 2 }

Release 1.0® (ISSN 1047-935X) is published monthly except for a combined July/August issue by CNET Networks, 104 Fifth Avenue, New York, NY 10011-6987; 1 (212) 924-8800; fax, 1 (212) 924-0240; www.release1-0.com. It covers the worlds of information technology and the Internet, including wireless communications, security, business models, online services, tracking systems, identity management and other unpredictable topics. . . and the policy issues they raise.

EDITOR: Esther Dyson
(edyson@edventure.com)

PUBLISHER: Daphne Kis
(daphne@edventure.com)

MANAGING EDITOR: Christina Koukkos
(christina@edventure.com)

CONTRIBUTING WRITERS: Dan Farber
(dan.farber@cnet.com), Dan
Gillmor (dan@gillmor.com),
Steven Johnson (stevenberlin-johnson@earthlink.net),
Clay Shirky (clay@shirky.com),
Dave Weinberger
(self@evident.com)

CIRCULATION MANAGER: Brodie Crawford
(brodie@edventure.com)

SYSTEMS MANAGER: Geoff Clarke
(geoff@edventure.com)

EDITORIAL COORDINATOR: Kate Tobin
(kate@edventure.com)

CONSULTING EDITOR: Bill Kutik
(bill@kutik.com)

Copyright © 2004, CNET Networks, Inc. All rights reserved. No material in this publication may be reproduced without written permission; however, we gladly arrange for reprints, bulk orders or site licenses. Subscriptions cost \$795 per year in the US, Canada and Mexico; \$850 overseas.

- In the physical world, a fruit can hang from only one branch. In the digital world, objects can easily be classified in dozens or even hundreds of different categories.
- In the real world, multiple people use any one tree. In the digital world, there can be a different tree for each person.
- In the real world, the person who owns the information generally also owns and controls the tree that organizes that information. In the digital world, users can control the organization of information owned by others.
(Exception to the rule: Westlaw owns the standard organization of case law even though the case law itself is in the public domain.)

These differences are so substantial that we can think of intellectual order as entering a third age. In the first, we organized the things themselves: We put books on shelves and silverware into drawers. In the second, we physically separated the metadata from the data: We built card catalogs and drew diagrams. In the third, the data and the metadata are digital, untying organization from the strictures of the physical world. In response, we are rapidly inventing new principles and tools of organization. When it comes to innovation on the Internet, metadata is becoming the new content.

But traditional taxonomic trees aren't something we can throw away without a thought. They are an amazingly efficient way of organizing complexity because they enable us to focus on one aspect (e.g., that's an apple) while keeping a universe of context (it's a fruit, part of a plant, a type of living thing) in the background, ready for access. Tree structures are built into our institutions. They may even be built into our genes. So we are in a confusing and fertile period as we try to sort out what works and what doesn't. Without trees, how would we organize college curricula, business org charts, the local library, and the order of species? How will we organize knowledge itself?

We may be on the path to finding out.

Webogeny recapitulates ontogeny

The tree of knowledge has roots, of course. They go back to Aristotle, who figured out how knowledge could be nested without having to claim that the container (say, the concept of human-ness) is the same sort of thing as what it contains (all existing humans). The individual items in a hierarchy *inherit* the properties of all the categories above it, so that if you know that Alcibiades is a human, you also know that he is a mammal and an animal. Inheritance provides a context by which the individual accretes the accumulated wisdom of the tree just by hanging on a particular branch – an amazingly efficient way of expressing knowledge.

Five hundred years later the Syrian philosopher Porphyry first drew Aristotle's system of nested concepts as a tree. That notion stuck, implicitly endorsed by Carl Linnaeus and Charles Darwin in the sciences, Francis Bacon in philosophy, and by libraries and academic departments just about everywhere.

The next stop in this story is Postmodernism's insistence that trees of knowledge are reflections of particular cultural assumptions and, importantly, conflate knowledge and power. You can't read Michel Foucault's *The Order of Things* and believe that order itself has no history. And not just French philosophers have given up on the old dream of finding a single, universal, comprehensive way of organizing the world's knowledge. You can't come out of Geoffrey C. Bowker and Susan Leigh Star's study of the International Classification of Diseases, *Sorting Things Out*, thinking that classification systems are value-free and objectively true. Nor can you look at the US Census' 2000 decision to expand the number of possible races without seeing that taxonomies can have enormous political and budgetary consequences.

The brief history of the Web has recapitulated Western culture's ontogeny of trees. Yahoo!'s directory tree became the early center of the Web, each leaf hand-selected and placed into categories designed initially by two computer science grad students at Stanford. But text search engines – AltaVista, HotBot, Google – dethroned Yahoo! as the Monarch of Search, and Yahoo! in turn has moved its browsable tree below the fold on its home page.

When text search isn't the right solution – for example, at e-commerce sites where people may not know the names of the products they're looking for – a more dynamic way of creating and presenting trees, called faceted classification, is coming into its own. Invented in the early 1930s by Shiyali Ranganathan, an Indian librarian, it applies a pre-defined set of parameters (or facets) to its objects. For example, watches might have facets such as manufacturer, digital or analog, men's or women's,

price, and electric or spring-driven. Some facets are a set of possible values (such as a pick-list of available manufacturers); others are a range of numerical values (such as price range). Users can then browse by selecting first on, say, digital or analog and then by price, or first by price and then by men's or women's. Users can drill down as they do with a normal tree, but the arrangement of the branches is dynamic and reflects the users' interests, not the store's. The store may not like it that you've routed around the \$25,000 Rolex they're offering on sale for a mere \$24,000, but you've found your \$50, waterproof, analog watch much faster.

Faceted classification still presents users with a hierarchical tree, making it easy for them to browse to what they want. But unlike traditional trees, faceted systems don't decide beforehand how the branches are arranged. For example, if an ice cream stand organized its "customer experience" around a traditional hierarchical taxonomy – a tree – it might have a customer first choose between two flavors, then among three sizes, and finally between a cup or cone. There are 12 potential paths and exactly one path to a large cup of chocolate ice cream. In a faceted system, you could browse first by flavor, size, or container, resulting in 36 potential paths and three ways of getting to your large cup of chocolate. Faceted systems, like trees, enable users to navigate by continually focusing their interests, but users get to decide how their interests are structured. This makes faceted systems very useful where there are lots of items with easily specifiable properties and users whose ways of browsing are difficult to predict, such as a parts catalog.

The long tail of tags

Tags have become the meme of the year, at least so far, writing another chapter in the history of classification systems. Tagging is an old idea, but it seems to be taking off now because some applications provide end-users with immediate benefits. For example, at del.icio.us, users enter bookmarks (URLs) they want to remember, adding a word or two – tags – so they can sort them later. Del.icio.us users can see not only everyone else's bookmarks, but also all the bookmarks tagged with a particular word. For example, if you care about Emily Dickinson, you can see all the Web pages del.icio.us users have tagged with "Dickinson" or "Emily Dickinson," a great tool for researchers.

Traditionally, people have been loath to attach metadata to objects, because it felt like a chore without immediate benefit. At del.icio.us and other sites such as Flickr, a photo-sharing site, there is a strong social benefit to tagging: We get to contribute to, and benefit from, the tagging done by others. To lower the hurdle and encourage

tagging, both sites allow us to type in any word we want, rather than forcing us to navigate some hierarchical, controlled vocabulary. Of course, that also makes it far harder to find relevant objects: There's no immediate way to tell whether a photo tagged with "apple" shows a fruit or a computer. Plus, a search for photos tagged with "apple" will miss relevant photos tagged as "GrannySmith."

Tags are a break from previous ways of categorizing. Both trees and faceted systems specify the categories, or facets, ahead of time. They both present users with tree-like structures for navigation, letting us climb down branches to get to the leaf we're looking for. Tagging instead creates piles of leaves in the hope that someone will figure out ways of putting them to use – perhaps by hanging them on trees, but perhaps creating other useful ways of sorting, categorizing and arranging them.

Even in these early days of tagging, we're seeing self-organizing taxonomies emerge from the piles. For example, if you're tagging a page about an Apple computer, you may notice that far more people use the tag "Mac" than "Macintosh." So, if you want lots of people to find the page, you will tag it "Mac." By using that tag, you have also increased the popularity and momentum of the "Mac" tag. The resulting bottom-up clusters of tags has been called a *folksonomy*. (It's also been called a "tagsonomy," but that's harder to differentiate from "taxonomy" when spoken aloud.)

Folksonomies stand in sharp contrast to both trees and faceted systems. First, folksonomies tend to be clusters of tags, not hierarchies: There's a pile of "apple" tags and another pile of "GrannySmith" tags, but the folksonomy may not recognize that the latter is a subset of the former. Hierarchies can sometimes be derived from folksonomies, but they don't have to be. Second, trees and faceted systems are designed ahead of time, usually by information professionals. Folksonomies grow organically. Third, trees and faceted systems are usually owned and controlled by the people who own the information being organized, whereas folksonomies are (so far) unowned and not centrally controlled. Fourth, trees and faceted systems drive out ambiguity. For example, take a page that in a tagging system carries the ambiguous tag "apple." In a tree or faceted system, the branch it hangs from would tell you whether the page is about computers or fruit – inheritance at work. Tagging systems are inherently ambiguous. Trees are neat; piles of leaves are messy.

Because of these differences, the three approaches are useful in different circumstances:

- Because they are unambiguous, trees work well where information can be sharply delineated and is centrally controlled. Users are accustomed to browsing trees, so little or no end-user training is required. But trees are expensive to build and maintain and require the user to understand the subject area well: How do you find the recipe for bread soup if you don't know to look in the "Tuscan Cooking" category?
- Faceted systems work splendidly where an application is being used by such a wide range of users that no one tree is going to match everyone's way of thinking. They are also easier to maintain than trees because adding a new item requires only filling in the information about the facets, rather than having to make a decision about exactly which category it should go into.
- Tagging systems are possible only if people are motivated to do more of the work themselves, for individual and/or social reasons. They are necessarily sloppy systems, so if it's crucial to find each and every object that has to do with, say, apples, tagging won't work. But for an inexpensive, easy way of using the wisdom of the crowd to make resources visible and sortable, there's nothing like tags.

The craft of creating and maintaining trees and faceted systems is well advanced and well understood. Businesses have been built around them. But we don't yet know the outcome of the current infatuation with tags. The potential is real: If tag-mania continues, it will provide a layer of new metadata, generated by humans for other humans, that will invoke innovation and businesses – and problems – we necessarily cannot anticipate.

The Stand of Trees

Trees – hierarchical taxonomies – don't have to be visible to be useful. In large, complex environments, trees can be cumbersome if presented to the user as such: Too many branches to walk down and only one way to get to any particular leaf. But trees have tremendous power. Not only do they make it easy to find other objects like the one you're looking for – if you get to the "motorcycle" category, you can also see "scooters" and "mopeds" – but they embody a schematic of thought that can be used to disambiguate search queries: An application can ask the user (or perhaps guess

based on other information) whether she is looking for “enterprise” as in “business” or “enterprise” as in the starship. That’s why trees are still hard at work, especially in organizations that have lots of data that doesn’t change very often. Where trees work, they work well.

Dewey Decimal Classification system: Power of incumbency

If you grew up in the United States, you were trained as a taxonomist. It probably happened in the third grade when you were marched down to the library and instructed in the ways of the Dewey Decimal Classification (DDC) system. In use in 200,000 libraries around the world, including 95 percent of US public schools, it is the archetype of a tree of knowledge. And that is exactly its problem. The very attributes that originally made it such an advance in the organization of physical books get in the way of its utility in the digital age.

Created in 1876 by the 23-year-old Melvil Dewey, about 15 million content items (books, videos, Web resources and more) now have DDC numbers and the Library of Congress alone adds 110,000 per year. Dewey, fresh out of a tiny, traditional Christian college, elaborated on Sir Francis Bacon’s division of knowledge into philosophy, history and art, adding six more top-level subject areas. He then divided each of these into ten further parts and continued his base-ten divisions to the right of the decimal point. This enabled books to be clustered on shelves by topic rather than alphabetically by author or title, as was common before Dewey came along. The geography of the library became a living map of knowledge through which we could walk and browse.

A hundred and thirty years later, the drawbacks of Dewey’s approach are readily apparent in our more diverse and tolerant society. The Online Computer Library Center (OCLC), the formal owner of the DDC since 1988 headquartered in Dublin, OH, and in the Washington, DC, offices of the Library of Congress, stresses the amount of work it has put into updating the classifications. Joan Mitchell, DDC editor-in-chief, gives an example of the sort of continuous modification the DDC undergoes: “It used to be that the religion section was almost entirely Christian. We’ve been working really hard across two editions to change that. We made a major expansion to Judaism and Islam.” That expansion, however, still gives each of those two religions just one integer among the 100 available for religious topics.

So why in this day and age is the DDC so biased towards Christians? First, changing DDC numbers requires sending legions of librarians armed with razor blades into

the stacks to scrape off the old numbers; the physical world is just plain hard to modify. Second, while to the right of the decimal point there's an infinite amount of space, there are only 1000 available integers, and integer real estate is worth more than fractional plots. There is no way for the DDC's ten-member editorial board to come up with revisions that will make the categorization scheme represent everyone's values, because we don't share all the same values.

The problem is compounded by the DDC's requirement to assign only one primary number to each book. While cross-references are allowed and electronic systems built on top of the DDC permit multiple filings, in physical libraries the DDC number is used to position books on shelves, and the laws of physics say a book can be in only one place at one time. That rule, of course, is routinely violated in the world of bits: At Amazon, *The Oldest Cuisine in the World: Cooking in Mesopotamia*, is classified under three categories: Gastronomy, Ancient Assyria, Babylonia & Sumer, and Middle Eastern History.

The DDC has 130 years of careful thought behind it and a worldwide body of people who are used to it. And Mitchell says, "I'm very interested in looking at how we can. . . find a way to make it [the DDC] a really useful tool on the Web. It could be an underlying tool" – providing identifying numbers presented in any taxonomical order one might want. Yet it's an unlikely candidate to provide the categorization schema for the Web because it remains a top-down structure for understanding everything, and everything isn't what it used to be: Not only do we disagree about what should go where, in the globally connected world, we know that we disagree about it. The DDC is, as Mitchell says, "a general knowledge organization system," but every day we discover – for better or worse – just how local the structure of knowledge is. The strengths of the DDC – its universality and stability – make it a tough sell to the ever-changing networked world.

Yahoo!: The inner tree

In the mid '90s, Yahoo!'s home-grown taxonomic tree was the magnetic north for Web searching, making the Web usable for millions of people. Now the Yahoo! tree has been pushed below the fold on its home page by a smorgasbord of services, ads and links. But, according to Yahoo!'s editor-in-chief Srinija Srinivasan – employee No. 5 at the company, with a background in cognitive science – that doesn't mean the tree is being chopped down. "The desire to browse never goes away," she says. Besides, "Even if we never showed the directory, we would still have built it because it continues to be our internal collective memory, our way of expressing, recording

and documenting what we know.” (This is a value the Dewey Decimal System has brought to the world of physical information resources.) “What we know” includes where millions of valuable Web pages live, what they’re about, and which other pages talk about the same topics; that knowledge has value even if these days Yahoo!’s users are locating those pages by searching for text more often than by browsing the tree.

The Yahoo! directory originally was built painstakingly by hand, and that is still the case. Now, however, it’s not merely the product of two Stanford computer science students – Jerry Yang and David Filo – who were trying to organize their own book-marks. Srinivasan heads a team of editors who select what to add to the directory, where it should go, and how it should be described. She won’t disclose the size of the team but says it’s still growing even as the home page has focused more on providing text searching. Says Srinivasan, “Even when we started, I’d frequently tell my team that the goal isn’t to have the be-all and end-all of classification systems. Our job is to know the Web, know what searchers want, and marry the two.”

The directory continues to help even those who are using other Yahoo! services, primarily by adding context. For example, suppose you use Yahoo!’s full-text search engine to look for pages about Iraq. If any of the pages that are retrieved are also included in Yahoo!’s directory, the search results page draws its descriptions from the one hand-created by a directory editor. Likewise, the search results page shows related information based on what else is in that page’s category. “We infuse the search results page with the context” drawn from the directory, says Srinivasan. “You don’t have to come to the directory. We can bring it to where you are.”

Yahoo! is committed to the value of its taxonomy as a data structure, even if people are now more comfortable with the search paradigm. “The extent to which you see the tree on the front page has no correlation to the amount of time we put into it,” Srinivasan says. “We have not let up one bit in seeing what human involvement can bring to information discovery online.”

Corbis: Controlled words and pictures

Yahoo!’s taxonomic tree is designed to allow users from everywhere on the Net find the best of what’s on the Net. Corbis serves a narrower group of people with a narrower set of objects: digital images. Further, Corbis has a strong economic interest in making sure users find every photograph that could conceivably meet their needs, so it has a team of nine fulltime cataloguers who categorize each image Corbis owns or

represents. “If a cataloguer looks at a photo of a beach scene and says beach, coast, coastal, and marine, and someone comes to our site and types in ‘seaside,’ that person won’t find the image unless we can automatically equate ‘seaside’ with those other terms,” says Joel Summerlin, search vocabulary manager at Corbis. “That’s why we broaden the side of the barn.”

Maintaining such a system is a serious undertaking. When a new image comes into the collection, one of the cataloguers uses home-brewed software to browse the 61,000 “preferred terms” in the Corbis thesaurus, or controlled vocabulary, for those that best describe the content of the image, typically attaching ten to 30 terms to each image. Since Corbis’ customers find images by typing terms into a search box

rather than by browsing the thesaurus, the system incorporates about 33,000 synonyms as well as more than 500,000 names of people, movies, works of art, places and more. That broadens the side of the barn so wide that if you misspell Katharine Hepburn’s name as Katherine or Catherine, you’ll still find most of the images of the high-cheek-boned screen legend in the Corbis collection. And if you’re looking for Moammar Khadaffi, at least 17 different ways of spelling his name will get you what you want.

CORBIS INFO
Headquarters: Seattle, WA
Founded: early 1989
Employees: 1000
Funding: privately owned and funded by Bill Gates
Key metric: represents more than 75 million images; \$170 million in 2004 revenues
URL: www.corbis.com

But why make this thesaurus hierarchical? It’s a big job to maintain it and it is fraught with the potential for embarrassment. For example, a few years ago, when you looked for images of “servants,” you’d find photos of “housewives” because, Summerlin explains, “Some

hyper-literal vocabulary editor decided that ‘housewives’ were a species of ‘domestic worker.’” Despite the occasional glitch, there are powerful advantages to maintaining a hierarchical thesaurus, he says:

First, if a person searches on a keyword, the results include its synonyms and narrower concepts. For example, a search for “dog” will return all different breeds of dogs, even if those images don’t actually have the keyword “dog” attached to them.

Second, a cataloguer only has to tag an image as “terrier,” and the system will automatically recognize it as a type of dog.

Third, “it helps us control homographs – words written the same way with different meanings,” says Summerlin. “We can pop up a box and ask you if you meant turkey the country, the bird or the meat.” Each of those three turkeys is a separate keyword, in a separate hierarchy.

Fourth, Corbis frequently acquires entire collections and represents images from other collections. Having a hierarchy enables Summerlin's group to blend the new collection's metadata with their own. "Let's say another provider has 100 images of wolves," he says. "They remembered to tag 70 with both 'wolf' and 'wolves,' but 30 only have the 'wolf' tag. By mapping them to our controlled term 'wolf,' which has 'wolves' as a synonym, we've improved their metadata. And we've added 'mammal' and 'canine' to boot."

Overall, Summerlin says, having a taxonomy "lessens the amount of superhuman effort cataloguers have to put into the system," and "gives you a system that you can alter or evolve as the needs and habits of your customers evolves." He adds, "I don't believe there's an über-taxonomy out there that will work for every possible scenario. At Corbis I've learned that your taxonomy has to be intimately bound up with your system and your users."

ClearForest grows branches

"We've found that a taxonomy is only effective as an information retrieval tool if it's pretty much consistent with how the end-user thinks," says Barak Pridor, CEO of ClearForest, a Waltham, MA-based company that finds data in business prose. Unfortunately, when the information is unpredictable – as is the case with most of what human beings write – it can be extraordinarily difficult for a software program to anticipate how a user will understand it. It's one thing, says Pridor, to drop data into an established taxonomy by noticing SIC codes in a form. It's another to figure out programmatically how to extract the entities, facts and relationships from a news article in a way that enables it to be categorized. For that, ClearForest deploys a combination of semantic analytic tools and domain-specific heuristics.

Pridor had been working with Technomatics, an Israeli company in the computer-aided production engineering field, when he met co-founders Ronen Feldman and Jonathan Aumann, both senior lecturers at Bar-Ilan University. "It was amazing to me to talk with two academics and hear that something didn't have to be 100 percent accurate to be useful." Pridor adds affectionately, "Ronen has the attention span of a fly," which perhaps explains why he built a system to help people focus on reading only what truly needs to be read.

Information is too surprising to be capable of exhaustive categorization; something unanticipated is always going to pop up. Automatic categorization systems that populate existing taxonomies may force novel information into inappropriate cate-

CLEARFOREST INFO
Headquarters: Waltham, MA
Founded: May 1998
Employees: 85
Funding: \$31.25 million from Pitango, Booz Allen, Harbourfest & Walden, Deutsche Bank Capital, Greylock Partners and others
Key metric: used by 30 client sites including Boeing, Dow Chemical, Eastman Chemical, EDS, and Thomson Financial
URL: www.clearforest.com

gories. Pridor points to Dow Chemical’s acquisition of Union Carbide in 2000. Dow not only acquired Union Carbide, it acquired all of the ways UC refers to chemical compounds. To have this deal accounted for by the SEC as a “pooling of interests” transaction, Dow Chemical and Union Carbide needed to merge their intellectual assets within 24 months. It was a daunting problem: Union Carbide had more than 100,000 documents, some dating back to World War II, that mentioned chemical compounds. Dow has four different hierarchical registries of chemical compounds, and Union Carbide had two of its own. Certain chemicals have more than ten different synonyms. In some cases there were seven names at Dow, five at Union Carbide, and only two that overlapped.

Dow used ClearForest software to go through the scanned documents, identifying ones that refer to chemicals. Experts then added them to the registry appropriately. Pridor emphasizes that this could not have worked if they had had to stick with an already-existing hierarchy. “It was a process of discovery,” he says, so the existing hierarchy wasn’t enough.

In fact, Pridor points to an example of a taxonomy getting in the way of seeing what’s there: “The property insurance business has an extensive code hierarchy for different property damage elements. It turns out that the number one problem in property insurance is mold. They didn’t have a code for it. It didn’t exist in the hierarchy. So it took the insurance business 20 years to discover that mold is their greatest problem.”

From Trees to Leaves

When Shiyali Ranganathan invented the Colon Classification System in 1933, he intended it as a new way of organizing books in libraries. The idea was both brilliant and eccentric. Ranganathan came up with five parameters (or “facets”) by which every book could be classified – personality, matter, energy, space and time – and a notation system to express that classification. (The parameters were to be separated by colons, explaining the system’s unfortunate name.)

While the Colon system didn’t spread much outside of India, the basic idea is radical and far-reaching. Rather than attempt to create a complete set of pre-defined subjects into which books can be pigeonholed, librarians using the Colon system can

combine the five facets to create new subjects as needed. That way, books don't have to be stuffed into a set of categories that could not anticipate how knowledge would develop.

The Colon system is not being applied on the Web, but the concept of facets in the frictionless digital world are the basis for highly flexible systems for browsing complex data. They allow users to traverse branches in the order that suits them, while providing clear, unambiguous access to what may be huge collections of items.

The Open Source Application Foundation's Chandler: Facets at work

Mimi Yin has performed with dance troupes in New York City and done some choreography. Now she's finding that faceted classification is a good way to choreograph users' interactions with the messiest and most voluminous personal data around: e-mails, contacts, tasks and schedules.

Yin is the user interface designer for Chandler, Mitch Kapor's open-source personal information manager (SEE **RELEASE 1.0**, JUNE 2004). She comes to the project after working on Web design and architecture for companies such as Roxio and MSN. Now she plans on putting a straightforward faceted classification system on people's desktops sometime in 2006, when Chandler ships.

Chandler had a "brief flirtation" with using trees before Yin got there, she says. "There are two levels at which we find trees problematic," she explains. First, "the notion that there's one single tree that can capture all of your knowledge is flawed to begin with." Second, "They show only one dimension: This is inside of that."

That's not enough information: "Is A inside of B because A is smaller than B, like the Russian dolls? Is it because the concept is smaller, the way 'lion' is inside 'animal'? Is it because A is less important than B?"

Third, the structure of a tree doesn't always represent the way we want to work with the data. "The pieces at the bottom of the tree aren't always the least important, although they do take the most clicks to get to. That's why we have desktop aliases [shortcuts]." And even if a tree does represent your interests, those interests change in the course of a day. For example, her research shows that people often start out organizing by projects but then want to see everything they have to do today. "Most people are just trying to keep on top of their lives and get their stuff done," she says. "It's hard for peo-

OSAF FOUNDATION INFO
Headquarters: San Francisco, CA
Founded: May 2001
Employees: 24
Funding: \$7.85 million from Mitchell Kapor, the Andrew W. Mellon Foundation, 25 universities comprising the Common Solutions Group and donations from individuals
Key metric: 0.5 release (pre-alpha) scheduled for March 2005
URL: www.osafoundation.org

ple to figure out whether ‘project’ should be a sub-node of ‘status’ and if ‘status’ should be a super-node of ‘time.’” There are reasons people need to go to school to become taxonomists.

So Chandler will use a faceted system – the full implementation of which will ship sometime after the program’s first release – to get over the limitations of trees, laying out its information horizontally in a grid that lets users decide with a click which facet will be the root and which will be the branches. For example, if the columns, left to right, are Projects, Teams, and People, the table will arrange itself into projects, each sorted into their various teams, and each team broken down into people. Drag People to the leftmost position and now the table will show you a list of people and the projects each is involved with. (Yin points out that this is similar to how iTunes works, except that iTunes always has the artist as the primary category.)

A PIM is a perfect place to apply a faceted system, for there are multiple parameters each of which the user might want to use as the root. And since the P in PIM stands for “personal,” it makes sense for the system to present information the way the user wants it at any particular moment. In the physical world, we’re stuck with static address books and file folders. In the digital world, letting users dynamically build their own trees is common sense.

Endeca: Facing the facets

No matter how overwhelming our e-mail and scheduling information looks to us, it’s a bouquet of daisies compared with the informational jungles large organizations quickly grow. Faceted classification scales up quite nicely, but getting it right requires taxonomic and content expertise as well as software that can handle huge computational problems fast enough to keep up with a user casually clicking through a series of screens.

“For the first year, when Endeca was under wraps, its working name was Optigrab,” says founder and CEO Steve Papa. “But when a prospect noticed that we’d named it after the little handle on eyeglasses invented by a Steve Martin character, we changed it to Endeca.” The name refers to “entdecken,” German for “discover,” a word that’s not only appropriate but is also free of any reference to the movie *The Jerk*. Endeca is now doing big business – doubling revenues year to year, and having its first \$10-million quarter – with organizations such as IBM, Wal-Mart, Barnes & Noble and the Library of Congress.

Faceted systems like the ones Endeca creates may *look* like the parametric searches featured on many e-commerce sites, but they're different. For example, at electronics retailer NewEgg.com, users shopping for digital cameras can specify the details of any of 13 parameters, from manufacturer to the type of memory stick, and see only the cameras that match those selections. But that's not yet faceted classification, explains Papa. The NewEgg system lets users specify the parameters to find all Nikon, 5-megapixel cameras for under \$50, even though there are no results. Endeca's system, on the other hand, dynamically adjusts the parameters so that users are never given choices that lead to null result sets. The result is what Endeca calls "guided navigation." Papa points to a demo Endeca constructed in-house using 90,000 reviews from the *Wine Spectator* database. Each review can be sorted on any of nine facets, including the type of wine, country, price range, year and winery. So far, it sounds like NewEgg. But if you say you want to see only wines with very high ratings, the checkboxes for the lower end of price range disappear, because there are no extremely good, cheap wines. Ask for a Zinfandel and all the countries except the United States and – surprise! – South Africa disappear. Faceted systems don't construct every logically possible tree, but only trees that can lead a user to a result she wants.

ENDECA INFO
Headquarters: Boston, MA
Founded: August 1999
Employees: 200
Funding: \$45 million from angels including Dick Parsons, Brian Totty, Bill Sahlman and institutions including Bessemer, Venrock, In-Q-Tel and Lehman Brothers
Key metric: \$32 million in 2004 revenues; key customers include The New York Times Company, the FBI, IBM, Wal-Mart, John Deere and the Library of Congress
URL: www.endeca.com

This is not a trivial technical challenge. Each time the user specifies a facet – "Show me white wines. Now show me white wines from Germany. . ." – the system computes the paths through the tree that result in populated branches. If a branch has no wines hanging from it, that branch doesn't sprout. In the wine demo, there are 1034 possible paths, but "only" 250 million of them lead to existing wines. That's a computable problem. If you were instead to hard-code a single tree structure with nine facets, there would be only 8,000 paths through the tree, and once you had sorted by country, you might not be able to sort by type.

These capabilities are crucial for an application Endeca built for a company that provides equipment and services to the oil and gas drilling industry. In one particular project, using a faceted user interface an engineer can find exactly the right bolt among 147,000 approved parts chosen from a database of 25 million available pieces, with a total of 1500 facets. The engineers need to see only the parts that are available, or else they will waste enormous amounts of time wandering in the desert. Further, because they see only facets relevant to the particular set of parts, they can browse a complex schema without having to know the schema itself.

Endeca is focusing on large customers with large online presences. For example, Barnes & Noble, which had hired eight ontologists to build a 250,000-term taxonomy, uses Endeca's system to provide users with results that cut across categories. Endeca's average deal price is about \$400,000, and the customers' results are frequently impressive, says Papa: Overstock.com increased conversion rates and revenue per session by double digits, because people were better able to find items; Eddie Bauer experienced a 30 percent increase in sales; and IBM.com saw a 50 percent increase in its conversion rate.

Endeca is now starting a push into the enterprise business intelligence space, using its faceted classification engine to produce dynamic reports. For example, Harvard University is rolling out a system that will enable about 1000 people working on alumni relations to sort donors by 20 different facets. Pick year, amount donated, and age facets and you are instantly shown a graph of donor data with year as the X axis, amount as the Y axis, and donor age groupings as the bars. Change a facet and the report updates to show the regions where a fundraising lunch is likely to be most lucrative. American Express, the US Army Reserve, Fidelity and NYTimes.com (SEE BOX, PAGE 17) are also customers.

With Endeca's e-commerce clients increasing their sales and its enterprise application clients gaining savings in the "tens of millions of dollars," according to Papa, Endeca is finding no shortage of work.

Siderean Software: Facets of trees

Siderean Software's name refers to sidereal navigation, the art of navigating the ocean by noting the rise and set points of 32 stars. The company literature contrasts its approach with giving the user nothing but a blank search box by which to navigate. The 32 stars become, for Siderean's users, whatever number of facets a project requires to make information findable. But even within Siderean's cutting-edge faceted categorizations, hierarchical taxonomies have not vanished: Open a facet and you may find a tree inside.

In pure faceted classification, all facets are equal: Pick any facet as your root, any other facet as the first branch, etc. That makes systems from companies such as Endeca excellent for navigating large, regular data sets such as parts catalogs. Siderean, on the other hand, is especially well suited for more complex schemas that involve complex relationships among the parts, such as a product catalog connected to a database of product developers, a sales team, and a library of technical papers.

NYTIMES.COM: ALL THE FACETS THAT ARE FIT TO PRINT

There may be 8 million stories in the Naked City, but there are 10 million in *The New York Times's* archives. The NYTimes.com wants to make sure you can find each and every one of them, all the way back to 1851, the year the *Times* was born. That means merging metadata that goes back more than 150 years and integrating multiple classification schemes and work processes. The site has to meet the needs of casual readers and dedicated researchers while also being comprehensible to search engine spiders. It's a problem set that's been bending the mind of Robert Larson, director of product management and development of NYTimes.com, for the past two years.

"We're in the process of completely overhauling our search infrastructure," says Larson, who previously was the information architect for the site. Casual users probably won't recognize that they're using an Endeca faceted classification system, because only one facet will be exposed on the initial results screen: date. Larson's research showed that when more facets were exposed, users tuned them out. Search for, say, "Senate" and the results will cluster themselves by year. But in an interface for advanced users, you'll also be able to drill down by other facets, if that's what your research calls for. With a faceted system, explains Larson, you can engage in "a type of horizontal searching." For example, you could see just the *Times's* editorials about the Senate, clustered by geography or by issue.

But the most important changes are occurring out of sight of the user. "We're building on the metadata that the Times Index department has been [generating manually] since 1851," says Larson. The tags comprise a controlled vocabulary with about 10,000 subject categories and many times that number of personal, organizational and place names. This is then used to build the big red Times Index volumes that are keyed to where the story is captured on microfilm. "About five years ago, we [the NYTimes.com website team] saw the value in all that metadata the Times Index had captured," says Larson,

realizing it could serve as the basis for a faceted system.

The site uses its volumes of metadata to create granular distribution feeds, to send Times News Tracker e-mail alerts and to target ads. "If an article is about the New York Yankees, we associate items from the New York Times Store with it so we can run contextual ads," says Larson. (It helps that the Store and the website use the same controlled vocabulary.) Conversely, "If we know that an article is about a tragedy, we make sure that no ads appear on that page." ("Tragedy" is not one of the subject headings; instead the system looks for tags such as "Disasters" or "Hate Crimes" that suggest it would be inappropriate to show advertising.) The system also uses the metadata to list related articles, i.e. articles with the same tags.

Most important, the metadata will allow the next version of the site, due in April, to create and maintain thousands of pages devoted to high-level topics. There will be topic pages for everything from Boston to Terrorism to Cloning. Topic pages will be like information dashboards surfacing the best content from the *Times's* various databases - reference books, news archive, photos, multimedia, discussion boards, etc. "This is the page the person who has a site about the Cuban Missile Crisis or Condoleezza Rice will link to," says Larson. At least as important, at last the search engines will have permanent URLs to spider instead of finding articles that in a few weeks are moved into the pay-per-view archives. Expect the topic pages to start showing up towards the top of Google search return lists within months after the site launches.

NYTimes.com plans to start hyperlinking between the day's news and these topic pages. According to Larson, "This allows us to tie the search experience to the news reader experience on the website. It will be a great service to readers who want to dig deeper and learn more about a particular subject." Not to mention, it will expose readers to more pay-to-read-more links to articles locked in the archive.

Building such systems benefits from human understanding and human effort in addition to transforming existing database schemas into facets automatically. (Less than 20 percent of Siderean's revenues come from consulting.)

SIDEREAN SOFTWARE INFO
Headquarters: El Segundo, CA
Founded: July 2001
Employees: 15
Funding: \$6 million from Clearstone, InnoCal and Red Rock
Key metric: customers include National Instruments, Environmental Health News and Fortunoff
URL: www.siderean.com

Founder and CTO Brad Allen was previously the CTO of Limbex, which created the consumer search assistant WebCompass. As an example of Siderean's value, he points to a project at NASA's Jet Propulsion Laboratory. After the Columbia space shuttle tragedy in 2003, it became imperative to break down the walls separating NASA's multiple sources of data. So the consulting company Taxonomy Strategies crawled the NASA data – structured and unstructured – and extracted metadata such as document type, originating organization or person and date. Under the supervision of Jayne Dutra, team leader for Web information architecture and Web content management at JPL, NASA then defined how it wanted users to be able to search the different facets: Abstracts and descrip-

tions should be full-text searchable, organizations should be searchable hierarchically, etc. Siderean represented those rules in an XML document from which its system builds HTML query boxes and menus for end-users to search and browse the documents. As a result, you can refine text searches by clicking on the appropriate facets listed to the left. For example, one specification of the "Missions and Projects" facet might be "Planetary Missions" under which would be listed the Apollo missions - but not the Mercury or Gemini missions, which were restricted to earth orbit.

Old-fashioned, pre-built hierarchies may surface during the search process. For example, says Allen, "If I'm narrowing down on the organization facet I might focus in on NASA center. Under that you would see a list of the different NASA centers. Click on one of those – the Wallops Flight Facility within the Goddard Space Flight Center, for example – and you get the next organizational level." He explains, "Facets can have information in them that is hierarchical or flat."

Adding the hierarchical information takes longer than simply setting up the faceted system, so Siderean sometimes delivers a purely faceted system first and then incrementally adds hierarchical elements. "Either or both can be effective in helping people focus their results," says Allen. Although the company occasionally goes head to head with Endeca, Allen does not see the companies as direct competitors. Much of Endeca's business comes from large e-commerce sites, while Siderean focuses on sales and marketing applications within retail, manufacturing and financial services, as well as applications within publishing, government and education. Less than 20 percent of its revenue comes from consulting. Both Endeca and Siderean will be presenting at PC Forum next month.

From Trees to Tags

Trees and faceted classification systems have two things in common. First, they both use established sets of categories to organize their contents, though trees nest the categories in one particular order while faceted systems allow the categories to be arranged dynamically. Second, both assume that there is an important difference between an information architect and a user: Architects design and users use.

Tagging systems violate both assumptions. They have no categories established ahead of time. Anyone can tag a resource with whatever text she wants, typically a word or two. Once resources have been tagged, applications can allow users to sort on them as they want. In fact, collections of tagged objects can be mined, clustered, or sorted into traditional taxonomic trees: Anyone can use the tagging metadata any way she can devise. Like faceted classifications, the user is in control of how the data is sorted; unlike faceted systems, the categories are not predetermined by the system's designer.

One particular type of ordering has grabbed the spotlight: bottom-up organization that arises from the aggregated behavior of the individual users doing the tagging. Folksonomies, as they are called, may emerge as users notice which tags are becoming popular, giving them an incentive to prefer those tags over others: If most people are tagging photos of the Statue of Liberty with the tag "StatueLiberty," if you tag it "FrenchGift," it won't be found as often.

Folksonomies may turn out to be another miracle of emergence. Or they could become a tyranny of the majority, a type of tagging colonialism. At the moment, all that we know is that a whole lot of tagging is going on. . .and that its own popularity will force tagging to evolve at Internet-speed.

del.icio.us: T.as.ty tags

The current tag-mania all started with del.icio.us. Although it's not the first site to let users tag objects, the site is showing people how valuable it is to socialize their tags.

The site's creator, Joshua Schachter (SEE **RELEASE 1.0**, JANUARY 2003), says simply, "Del.icio.us is bookmarks," but tagging was part of it from the beginning. At del.icio.us, users post links to pages they want to remember, optionally attaching a word or two as a tag to help them sort through the pages. A user's del.icio.us home page is similar to a list of bookmarks in a browser, but at del.icio.us a user's bookmarks and tags are visible to others. You can see all the other bookmarks others

have tagged with some word you're interested in. You can even subscribe to a tag and see its latest bookmarks in your RSS feed aggregator (SEE **RELEASE 1.0**, JULY 2003 AND DECEMBER 2004).

Schachter traces the site back to a moment in the late '90s when he realized he had 20,000 lines in the file of bookmarks he kept for himself. At that time he was maintaining Memepool, a site he created in 1998, which he describes as an early blog. "I came across lots and lots of links I wanted to write about," he says. "I had a file in my home directory on my hard drive and I just pasted links in it." Recently going through that file, he found the tag that started it all: "Eight lines into that file there's a URL, a space, a hash mark and then a tag: 'math.'" By adding tags, he was able to "grep" (search for) all bookmarks on a particular topic. So he built a site called Muxway – now defunct – that listed his bookmarks, complete with tags, mainly so he could easily point his friends at, say, his links that talk about WiFi. Schachter was looking for a new project and thought about doing something "midway between

blogs and Muxway," he says. "But then I hacked Muxway to make it multi-user, and instead of going halfway between Memepool and Muxway, a friend suggested I go halfway between Friendster and Muxway, and I ended up with del.icio.us." That was in 2003. The site now has 60,000 users and more than 1 million unique bookmarks, with an average of two tags per bookmark.

DEL.ICIO.US INFO
del.icio.us
Headquarters: New York, NY
Founded: September 2003
Employees: 0
Funding: undisclosed amount (pending)
Key metric: 1 million bookmarks, with an average of two tags each
URL: del.icio.us

The question is: What will happen when del.icio.us has 60 *million* users and half a billion bookmarks or links? How will social tagging hold up? Schachter isn't fazed by the prospect. "There are some statistical things I can do that the librarians don't understand and don't like," he says. Already del.icio.us suggests alternatives, drawn from other people's tags, once you put more than ten items under one tag.

For example, if you have ten pages tagged as "weblog," when you bookmark the 11th, del.icio.us can recommend other people's tags as being related to yours. This is not implemented at the moment because "My current CPU doesn't have enough cycles," Schachter explains. The system also currently favors "interestingness" over popularity, which Schachter describes as the first derivative of popularity – i.e., the change in popularity. For example, Google.com is so frequently bookmarked that it's not an interesting bookmark. On the other hand, Schachter says that about an hour after Steve Jobs' keynote at MacWorld last month, the Apple page about the new Mac Mini floated to the top of del.icio.us' list of interesting pages. "That's pretty good for a machine," he says.

Del.icio.us doesn't give much guidance to users about how to tag their pages. For example, when you bookmark a page, it doesn't tell you which tags have most frequently been applied to it. But that feature is on Schachter's development plate. (He works full-time in New York City in an unrelated financial job and works on del.icio.us in his free time.) He says the new user interface will show users the popular tags, but he stresses that it will show them below the list of your own tags. "I don't want people to be overly influenced by what other people are doing. Tiny, subtle human factors will influence people's behavior, so I'm trying to be very careful and follow my intuition, and sometimes that's a bit hard to hear, so I have to go slowly."

Besides, Schachter is not aiming to create a system of perfect tags. "Del.icio.us is an amplification system for your memory of URLs," and memory works best, he believes, by instinct. He wants to encourage people to tag things the way that makes sense to them in that first instant, not to go along with the most popular tag. "The problem with popularity is that it persists," he says. "Things are popular now because they were popular a bit ago. I want to scale down the influence gained by being popular." In fact, he dismisses the urge to harmonize tags as "the librarian instinct" – treating "blog," "blogs" and "weblog" as the same tag. "Someone's going to tell me what tags I ought to use?" he scoffs. "That trumps my intuition, and that's the most powerful thing you have going for you."

Schachter is exploring ways in which users could tag bookmarks idiosyncratically but still cluster them successfully by topic, even when del.icio.us grows to 60 million users. "Standard matching algorithms work well," he says, enabling del.icio.us to find other people who think like you. "At the same time, people who are not at all like you are also very interesting. I'm actually maximizing the similarity and the difference, which almost always gets you interesting stuff to read."

Schachter also wants to take advantage of group knowledge. "Right now there are two scales of use at del.icio.us: personal and site-wide. It needs groups. I'm working on it. I just have to sit down and write it."

These techniques will, he thinks, work pretty well but not perfectly. "The fact is that a URL can be 73 percent in a category. The edges are fuzzy." But that's fine, because del.icio.us does not aspire to be a perfect information resource. "The task is not categorizing things. The task is remembering in public so that other people can retrieve things in some manner."

Flickr: Photo tagging

Flickr, a site for posting and sharing photos, learned a lot from del.icio.us.

(DISCLOSURE: ESTHER DYSON IS AN INVESTOR.) Photos are very different from bookmarks: While many people at del.icio.us may bookmark the same page, each photo at Flickr is unique. Flickr, however, knows something about its users that del.icio.us doesn't: social groupings. The result is a fascinating intersection of tags and social networks. Further, Flickr is facing the scaling issues earlier than del.icio.us is, driving it to adopt a clustered view of tags – a harbinger that as tagging succeeds, it is going to get usefully lumpy.

“Albums break down as a way of organizing photos when [an individual] gets too many of them,” says Flickr co-founder Stewart Butterfield, a former designer, consultant and entrepreneur. Besides, pasting photos into a real-world album requires that you choose one primary parameter – usually date. “You should be able to cut

through them orthogonally,” says Butterfield. By using multiple tags, we can see our photos not just chronologically but sorted by the people who are in them, who posted them, their locations, the events they record. . . whatever tags any user has added. Further, you can see all the public photos that Flickr members have tagged with a particular string. These aggregations-by-tag are available as RSS feeds; Butterfield reports that the top three are Graffiti, Decay and London. Why “decay”? “Beautiful photos,” he replies.

While it's fun to see the photos the worldwide Flickr community has tagged, the photos your friends posted have special meaning. So Flickr added a social-networking capability: You can declare people as contacts or family – de-facto private groups – see what they've posted (tantamount to bookmarking their photo streams), and even allow them to tag your photos, thus distributing a task that, done alone, just about no one enjoys. Done as a group for someone else, it

can be an enjoyable form of social engagement. (Esther Dyson has commented that tagging a friend's photo reminds her of how her brother's pet monkey used to sit on his shoulder and poke through his hair for lice.)

The system is simple and, more important, fun. Since the site went live as a photo site in June 2004, it has grown to 308,000 users and 4.5 million photos with 5 million tags. An impressive 64 percent of photos have at least one tag. That's a lot of tags, and a lot of social interaction. Butterfield says Flickr is looking at various steps to help make tag navigation easier.

FLICKR INFO
Headquarters: Vancouver, Canada
Founded: June 2002
Employees: 11
Funding: \$1.4 million from angels including Reid Hoffman, James Currier, Brett Bullington, Esther Dyson, and friends and family
Key metric: of 308,000 registered users, 56 percent have added at least one tag; of 4.5 million photos, 64 percent have at least one tag
URL: www.flickr.com

Already, if you search for a popular tag, Flickr will show you related tags: A search for “Italy” causes Flickr to suggest Rome, Venice and Florence. “If there are a sufficient number of tags, the accuracy of the suggestions is uncanny,” says Butterfield. “And it’s done without human intervention.” Del.icio.us can discover these relationships by looking at how multiple people have tagged the same bookmark, but every image in Flickr is unique – presumably you’re the only one posting those photos of your trip – so it instead analyzes how tags cluster: Many photos tagged as “Italy” are also tagged as “Rome.” And if two people have tagged photos with “Italy,” the system also looks for patterns in the other tags they’ve used on other photos. “There’s a lot more of that coming,” says Butterfield. He promises to handle the privacy issues carefully. For example, clusters of tags have to be spread across many users to affect the system’s behavior, so no individual’s way of thinking gets exposed. (This should also help fight tag spam, a growing concern at every tagging site.)

Not all of the information is going to be gathered bottom-up. Some important relationships can be taught top-down. “Geographic relationships are particularly important,” says Butterfield, since so many photos are of places. “And definitely synonyms. And possibly opposites,” he says. He hopes to deal with synonyms, plural and singular forms, mapping across languages, etc., in a bottom-up way that, he says, “is crazy enough that it just might work.” Butterfield is not yet ready to get specific about it. He’s also looking at introducing a handful of special kinds of tags that would invite users to input specific information. For example, there might be a special tag that indicates that it refers to where the photo was taken, making it easier to find photos of, say, the state of Georgia without also seeing photos of people named Georgia.

By adding these features and capabilities, searches on tags at Flickr will return even more photos, which raises its own problem. “We’re rolling out ‘interestingness,’” says Butterfield, using the same term as Schachter at del.icio.us. “It’s like page rank for pictures.” It will rank photos by looking at a dozen usage characteristics, including how many times it was viewed, how many comments it drew, and how many times it was added as a favorite. “It counts more if a stranger takes it as a favorite than if someone in your contact list does,” Butterfield says.

Interestingness isn’t a perfect solution to the problem of there being just too many photos. Nor does it need to be, according to Butterfield: “It’s okay if we show only 80 percent or even 50 percent of the photos of the Bay Bridge in San Francisco because people have tagged it inconsistently,” he says. “No one needs to see them all.” This is the type of statement that drives traditional librarians mad. But then, Flickr wasn’t built for them.

Butterfield sees several revenue opportunities for Flickr. A premium service, already in beta, will offer unlimited storage, ad-free browsing, and other such benefits for \$60 per year. He is also considering charging for services such as printing and DVD backups and instituting context-based advertising – a page of photos of Italy might have ads for tours of Rome. He also sees an opportunity, smaller than that offered by advertising, to turn Flickr into a marketplace where people can sell their photos to the media and to other users. Flickr does not make its revenues public, but Butterfield says “we are making money,” enough to cover the costs of its infrastructure “many times over.”

Wikipedia: The shape of grass

Here’s a Zen koan: If a non-hierarchical group creates a hierarchical taxonomy, what does it look like? We can find one possible answer at Wikipedia, a not-for-profit online encyclopedia in which the entries are generated, written and edited by a self-selected group of volunteers. Wikipedia’s bottom-up approach to categorizing and taxonomizing has elements of folksonomy, but, as with much of the project, it mixes the spontaneity of the grassroots with rigorous community-based control.

Jimmy Wales, director of the non-profit Wikimedia Foundation, came up with the idea for Wikipedia when he was trying to find areas other than software development where an open source approach might work. “To me, the encyclopedia is an ideal candidate for this because it has a kind of objective nature that people can agree on,” he says. Since the site went public in January 2001, volunteers have added 1.3 million articles, including more than 450,000 in English. Users find articles by searching on the site – and increasingly, Wikipedia articles are showing up at the top of results at Google – as well as by following the many links within almost every article.

The articles are also categorized by topic, a late addition to the site. “Having categories was frequently opposed and there was a lot of brouhaha about how to do it,” says Wales. “We decided that it doesn’t make sense to have a formal top-down tree that has to be a certain way. Instead, the categories are in a sense tags.” This is in line with Wikipedia’s philosophy of doing nothing top-down unless it’s absolutely necessary to maintain the quality of the site. Anyone can create a category, place it as a sub-category of an existing one, and assign an article to multiple categories. About 75 percent of the English-language articles are categorized.

You can see the result – all the categories that have been invented so far – on the Wikipedia Categories page. The tree is quite flat, but Wales says it has some useful

features simply not found in top-down taxonomies. He points to the British Broadcasting Corporation's taxonomy of websites, similar to Yahoo! and DMOZ, an open source Yahoo-like Web directory. "They've worked diligently on this for years," he says. "They were amazed by our category system because it has levels of detail that they can't get to. We have a list of football [i.e. soccer] players by country. It goes on forever. It's the sort of thing they can't do with three or four employees because they don't have the local domain knowledge. But we have local people who are self-selected and know about the area." In other words, this is the type of detail you cannot achieve with just three to four employees but that you can get if you have exactly zero employees.

The category system is used most frequently as a way of exploring a topic more broadly once you've navigated to a particular article. For example, at the bottom of the John Coltrane article, seven categories are listed: 1926 births, 1967 deaths, jazz saxophonists, jazz composers, Miles Davis, United States musicians, and Philadelphia, PA. Click on "jazz saxophonists" and you'll get a list of 104 articles with that tag. Click on "United States musicians" and you'll find 189 articles and 11 sub-categories. The subcategories, however, are rather random: "California musicians" and "American guitarists," but also "Alice Cooper members" and "Tori Amos." Why is Tori Amos a category instead of a listing? It's not because some Amos fan insisted that she rise above the other 189 musicians in the "US musicians" category. Rather, someone decided it made sense to list "Tori Amos albums" and "Tori Amos songs" as sub-categories of "Tori Amos." Once Amos has a sub-category, she now counts as a top-level category at the same level as "American composers."

This is the type of inconsistency that drives professional taxonomists insane. Wikipedia leaves ironing it out to the wisdom of its masses of users; eventually someone will either re-tag the Tori Amos entries, or people will decide it makes sense to tag most musicians that way. If categorization were to go wildly wrong, there are approved volunteers who patrol Wikipedia at the tag level, straightening out such anomalies. To help them do their "job," Wikipedia now posts changes to tags on a separate "recent changes" page so that categorization edits don't get lost in the shuffle of typo fixes. But that also points out the essential difference between the del.icio.us and Wikipedia approaches to tagging. At del.icio.us, tags are personal to the users who apply them, and a popular page can have dozens of different tags applied by thousands of users. At Wikipedia, tags are more generic and are applied not so much by casual readers but by people

WIKIPEDIA INFO
Headquarters: St. Petersburg, FL
Founded: June 2003
Employees: 0 (12,000 active volunteers)
Funding: non-profit self-funded by founder and from donations and grants; \$500,000 2005 budget
Key metric: 400 million monthly page views
URL: www.wikipedia.org

acting as editors of the page. The editors tag the page to get it listed in the centralized Wikipedia categorization system; wacky or wildly idiosyncratic tags are likely to be edited out. (Of course, if you want to remember a Wikipedia page with your own personal tag, you can always bookmark it at del.icio.us.)

Currently, the categories are designed only to help users explore the site, Wales says. The category system does not yet help provide more relevant search results because the system's search is based on MySQL, which does not easily lend itself to integration with other metadata systems. Wales promises the site's upcoming move to a new search engine "will incorporate different algorithms, so we can look at the category system and see how we can use it as hints." He also would like recent changes in categories to be available as RSS feeds. "You could subscribe to the recent changes in, say, jazz musicians. You'd get a feed of every article that's changed in that category – which would be great for editors – or new articles in a category." But that's in the indefinite future.

In the meantime, Wales will continue to use the category system for personal fun. One of his favorite categories: "Fictional pigs, a sub-category of fictional animals. There are a surprising number," he says, rattling off a few, from Snowball in *Animal Farm* to Wilbur in *Charlotte's Web*. "I think that stuff is a hoot."

frassle: Same tag, different name

One of frassle's many taglines – or anti-taglines – is "Built out of duct tape and drinking straws." But that's just geek humility. It's actually a nights-and-weekends project for its two developers, Shimon Rura and Josh Ain, developers of workforce management software at Kronos, who became friends at Williams College. At its heart, frassle is an open-source blog-hosting site that treats blogs not just as content but as nodes in relationships. The relationships are subtle and shifting, so frassle pays particularly close attention to the metadata created by users as they use the tool. Chief among these are the taxonomies individuals build for themselves.

Frassle makes all blogging content instantly re-usable, so you can spin up a dynamic page containing bits from blogs and comments hosted by frassle or available via RSS. Frassle encourages tagging as one way to cluster content by enabling users to create their own hierarchical taxonomy of categories, like a foldering system; you apply a tag to a post by clicking on one of your categories. Rura explains that there is no default taxonomy in frassle, because the site respects its users' particular ways of organizing their thoughts and activities.

Enabling users to create their own taxonomies serves their need to find their own stuff quickly, Rura says, but it also helps the user find interesting stuff that others have written. The problem is that, without a top-down taxonomy or controlled vocabulary, two people may give different names to highly related categories: What you call “politics” is related to what I call “elections” and what someone else calls “partisan propaganda.” So frassle looks at the one stable, reliable element in this equation: URLs. Says Rura, “When a new link comes into the system within a post, frassle looks at all the other categories that contain that link and does a simple statistical calculation. If two categories contain some percentage of identical links, the system assumes they are related” – the same approach Schachter is looking at for clustering pages.

FRASSLE INFO
Headquarters: Arlington, MA
Founded: November 2003
Employees: 2 primary developers
Funding: self-funded
Key metric: 700 registered users
URL: www.frassle.net

Rura says, “If I had to look in my crystal ball and see the ultimate application of this technology, it would be personalized search. There’s lots of hubbub about it right now, but the missing piece is a model of the information a person is interested in.” Taxonomies based on users’ links or tags can provide that missing piece, especially when applied socially: Frassle can use the link between my “Election 2004” tag and others’ “politics” tag as a filter when I search for articles about politics, increasing the relevance of pages tagged “politics” and decreasing the relevance of, say, spam. And, Rura says, this could scale beyond your own social circle: “Google seems eerily well-prepared for this, thanks to its ownership of Blogger and Orkut.”

Rura says of frassle: “I don’t know what it’s really for. We think these are interesting problems to work on, and it’s wonderful when people get excited about it.” The site is currently in alpha, with no target date for moving into beta. “We welcome new users and developers,” says Rura.

Technorati: Searching for tags

While sites like del.icio.us, Flickr and frassle work on ways to let us search across disparate tags (the “US = U.S.A. = America” problem), the situation gets much more challenging when you try to pull together information across different sites. For that it would seem one needs a third-party tag broker. That’s the challenge Technorati stepped up to last month. (DISCLOSURE: ESTHER DYSON IS AN INVESTOR AND BOTH DYSON AND AUTHOR DAVID WEINBERGER ARE ON THE BOARD OF ADVISORS.)

Technorati (SEE **RELEASE 1.0**, JULY 2003, MARCH 2004 AND DECEMBER 2004) is a search site that indexes more than 7 million weblogs in near-real time. Although blogs are not tagged, many blogging software packages allow authors to create categories for their blogs so readers can, for example, see all of a blogger's posts about astrophysics and skip the ones about aromatherapy for cats. (Unlike frassle, most systems don't let users create hierarchical categories.) Technorati decided that when it full-text indexes a post, it will notice the categories and treat them as tags. If your software doesn't support categories, or if you want to add additional tags to your post – categories tend to be broader than tags – Technorati lets you use an extension of the standard linking syntax to do so.

Since Technorati was getting into the tagging game, founder and CEO Dave Sifry decided it could also pick up the tagged objects at Flickr and del.icio.us using those sites' APIs. It has since added tagged bookmarks from Furl.com, another bookmark site. As a result, when you search for a tag at Technorati, you're shown a dynamically-

created results page that lists the weblogs Technorati has indexed, bookmarks at del.icio.us and Furl and photos at Flickr, that share the tag. If you want to browse instead of search, Technorati has a page with the top 200 tags in alphabetical order, with the font size representing the number of tagged items.

TECHNORATI INFO
Headquarters: San Francisco, CA
Founded: November 2002
Employees: 16
Funding: undisclosed amount from Draper Fisher Jurvetson, Mobius Venture Partners and angels including Esther Dyson
Key metric: more than 1.5 million tagged posts and 150,000 tags were indexed in the first two weeks after tags were launched
URL: www.technorati.com

Sifry is hardly a tag nut. "I was very anti-tag for a long time," he says. "In 1993 and 1994, everyone in the search business was saying that people can put their keywords in so the search engines can find your articles better. It was great for about six months until the spammers came along." Now, Sifry says, "People are using categories and tags because it helps them to organize their lives. But as soon as you aggregate all of that information and make information available about the tags themselves, they become these emergent, self-organized things. We're allowing communities of interest to form based

on selfish interests. It's a virtuous circle." Among the unforeseen effects: A group of Chinese bloggers and a separate group of Irish bloggers are using Technorati tag pages as a group blog. "People are now using Technorati as this automatic mini-site, like a quick, lightweight portal that can be created about any topic," says Sifry.

Sifry is offering this service for free, with the intention of "monetizing" it via content syndication, advertising and sponsorship. Technorati itself began after Sifry left Linuxcare and eneded something to do; he started Technorati as a way to be of ser-

vice to the community – a lesson he and his brother Micah were taught by their politically active parents. (Micah is a left-leaning writer and activist.)

Pay services are on the way, including ones that mine the anonymized data Technorati gathers about blogs. And Sifry expects that Technorati's new position as a tag aggregator eventually will lead to revenues as well. "Tags have some pretty interesting sponsorship possibilities," he says. He's also open to syndicating the aggregated content to media outlets. "Those are two possible business models," he says, "but we're looking at where this goes. You can't do this altruistically unless you come up with something that's sustainable as well."

The Tagging Future

Hierarchical taxonomies have been under fire for decades, their limitations exposed by the scale, wildness and anti-authoritarianism of the Internet. We are right at the beginning of the tagging revolution. There are more questions than answers, as is only proper for a change affecting something as deep as how we classify our stuff. The questions fall into two categories: What's going to happen? And what effect will it have?

The easy answer to the first question is: Innovation will happen. Beyond that, it's possible to predict some obvious problems that will be addressed in unpredictable ways.

The biggest problem has to do with scaling. What happens when sites have not tens of thousands of users (del.icio.us) and not hundreds of thousands of users (Flickr), but millions and tens of millions of users? What happens when a search for the tag "San Francisco" returns some substantial fraction of the 26 million results that phrase currently gets at Google? When a tagging system gets big, how can it deliver relevant results?

Scaling also brings exactly the opposite problem: Not all of the relevant objects are tagged with any particular tag. So not only will I be overwhelmed with photos of San Francisco, I won't get the ones tagged "SF," "SanFran" or "Golden Gate Bridge." When what information architects call "recall" is the issue – finding each and every photo of San Francisco – there are a few obvious paths to explore: You could algorithmically analyze large patterns of tags and guess with some probability that "SF" counts, or that "SF" counts if there's also a photo in the same set labeled

“GoldenGate.” Or, you could build or buy a hierarchical gazetteer that knows that “Tenderloin” is a district of San Francisco. Or you could encourage people to use “SanFran” when they tag photos of San Francisco.

All such techniques attempt to add context back to the tag. After all, a tag consists of just a few letters with no further metadata built into it: Taken by itself, we can’t tell whether, say, “SF” refers to a city, is someone’s initials, or is the equivalent of “BS” in some obscure language. The simplicity of tagging accounts for much of its appeal, but is also its greatest weakness.

Elizabeth Lawley, a professor of information technology at Rochester Institute of Technology, worries that a purely bottom-up approach will not only de-contextualize tags but encourage people to use such broad and generic tags that they won’t be very useful. She points to The ESP Game, an experiment by Luis von Ahn and Laura Dabbish, professors at the Carnegie Mellon School of Computer Science. The site shows you an image and asks you to type in a word describing it. Another person somewhere else on the Net is paired up with you, performing the same task simultaneously. Your joint aim is to come up with the same word before time is up, without using any of the words shown in a “forbidden” list – presumably words already assigned to that image. Lawley observed that in her experience, the winning words tended to be context-free and superficial: A Greek coin got tagged as “round” rather than as “coin” or “Greek.” “I think,” she wrote on the Many2Many group blog (for which author David Weinberger also blogs) “that the same factors that influence players of the ESP Game to try to maximize agreement rather than depth are also at work in the new folksonomic playgrounds. Increasingly, people are changing the way they label their links or photos because of how they see other people labeling them.” That’s why encouraging people to use popular tags troubles Schachter, creator of del.icio.us.

COMING SOON

- PC Forum documentation
- genetics, testing and health
- identity and life on the Web
- And much more. . . (If you know of any good examples of the categories listed above, please let us know.)

So, we are on the horns of a dilemma. If people don’t use the same tags, we get fragmentation. If people do use the same tags, we might get dumbed-down, useless tags and a tyranny of the majority. If there are too many items tagged with a particular tag, there will be too many to find. If not all similar items are tagged with the same tag, we won’t find everything.

A little bit of context can go a long way when it comes to filtering and clustering tags. For example, if I know that a “web” tag was created by a spider fancier, I will cluster it differently than if it was cre-

ated by an Internet technologist. This makes the intersection of tagging systems and social networks especially promising. Social data such as the behaviors of your friends and people statistically like you may provide sufficient clues to blow relevant data to the top of the leaf pile. The utility of such implicit recommendation systems could provide an incentive to form social groups based on semantic similarities: By joining the Bruce Springsteen affinity group, my tagging application can figure out that when I want to see photos of The Boss, I don't want to see pictures of middle-aged, bald white guys smoking cigars.

Still, questions arise faster than answers: Will social groups agree on tag sets? For example, Ethan Zuckerman, at the Harvard Berkman Center for Internet and Society (AUTHOR DAVID WEINBERGER IS ALSO A FELLOW THERE), is suggesting that those involved in the GlobalVoices project – an initiative trying to increase the visibility of bloggers who provide insight into life in their countries – use the “gv-“ prefix when tagging pages about developing nations. Will formal tag sets be established for particular social groups or communities of interest? Will some standard fields be established for parameters that could help with many searches, such as “country,” “language,” and “author”? (The Dublin Core, a standard set of metadata categories to be attached to written works, has been proposing just such a development for years.)

Assuming we collectively address the scaling issues sufficiently that tagging remains useful and appealing, there are yet more challenges:

- Who owns tags?
- Who owns the way they could be pulled into relationships, creating a taxonomy or ontology?
- What do we do about tag spammers?
- How do we internationalize tagging?

There is a simple solution, however, to all of these issues: Create the tags and experiment. Tags are becoming a new layer of infrastructure. They will enable yet another round of creativity as we figure out, collectively, what variety of things we can do with this metadata.

And as we do so, we will inevitably build new businesses, with different business models. Some we can predict:

- Software companies sell tools to help end-users tag
- Software companies and systems integrators provide clustering tools to pull together objects tagged differently but with related content

- Content experts create tag sets, perhaps enhanced with schema of their relationships, and sell access to them to users, enterprises and industries. This could include thesauruses, gazetteers and controlled vocabularies to broaden results from what a tag says to what the tagger meant
- Tag brokers enable us to share tags and tagging schemes, perhaps on a subscription basis
- Data mining services, extracting value from the newly tagged Web, are provided for a fee by consulting companies.

The most important opportunities are, of course, the ones we can't predict. ■ **R 1.0**

Resources & Contact Information

Louis Rosenfeld, *Consultant*, 1 (734) 663-3323; lou@louisrosenfeld.com

Mimi Yin, *User Interface Designer, Chandler Project, Open Source Application Foundation*, 1 (415) 946-3012;
mimi@osafoundation.org

Barak Pridor, *CEO, ClearForest*, 1 (781) 250-4300; information@clearforest.com

Joel Summerlin, *Search Vocabulary Manager, Corbis*, 1 (800) 260-0444; joels@corbis.com

Joshua Schachter, *Founder, del.icio.us*, 1 (917) 670-6015; joshua@del.icio.us

Steve Papa, *Founder & CEO, Endeca*, 1 (617) 388-4138; spapa@endeca.com

Shimon Rura, *Founder, frassle*, 1 (857) 928-3028; shimon@rura.org

Josh Ain, *Developer, frassle*, 1 (617) 780-9949; josh.ain@gmail.com

Jayne Dutra, *Team Leader for Web Information Architecture and Web Content Management, Jet Propulsion Laboratory, NASA*, 1 (818) 354-6948

Stewart Butterfield, *President, Ludicorp Research & Development*, 1 (604) 551-8514; stewart@ludicorp.com

Robert Larson, *Director of Product Management and Development, NYTimes.com*, 1 (646) 698-8136;
robert@nytimes.com

Joan Mitchell, *Editor-in-chief, Dewey Decimal Classification, OCLC*, 1 (614) 764-6000; mitchelj@oclc.org

Elizabeth Lawley, *Director of the Lab for Social Computing, Rochester Institute of Technology*, 1 (585) 475-6896;
ell@mail.rit.edu

Brad Allen, *Founder & CTO, Siderean Software*, 1 (310) 647-4266; ballen@siderean.com

Dave Sifry, *Founder & CEO, Technorati*, 1 (415) 846-0232; dsifry@technorati.com

Jimmy Wales, *President, Wikimedia Foundation*, 1 (727) 231-0101; jwales@wikia.com

Srinija Srinivasan, *Vice President & Editor-in-chief, Yahoo!*, 1 (408) 349-3300; srinija@yahoo-inc.com

For further reading:

Many2Many, a collaboratively authored blog on social software: <http://www.corante.com/many>

Dublin Core metadata initiative: <http://dublincore.org>

Foucault, Michel, "The Order of Things: An Archaeology of Human Sciences," New York: Pantheon Books, 1970

Bowker, Geoffrey C. & Susan Leigh Star, "Sorting Things Out: Classification and Its Consequences," Boston: The MIT Press, 1999

Lakoff, George, "Women, Fire and Dangerous Things," Chicago: University of Chicago Press, 1990

Aristotle, "The Metaphysics," New York: Penguin Classics, 1999

Calendar of High-Tech Events

- MARCH 7-8** **Digital Living Room** - San Mateo, CA. The Digital Living Room summit addresses the technologies and services that are transforming the living room into a digital hub, including HDTV, digital video recorders, telecom, media center PCs, recordable DVDs, platforms, broadband Internet, ethernet and WiFi, video-on-demand, multiplayer and next-generation gaming, mobile and wireless, converged devices, streaming and downloaded music and video, digital rights management and much more. Speakers will include Walter Mossberg (Wall Street Journal), Yari Landau (Sony Pictures Digital) and Rob Glaser (RealNetworks). Register via the site, or contact iHollywoodForum's, 1 (310) 815-3884 or info@iHollywoodForum.com. www.digitallivingroom.com
- MARCH 7-10** **Semantic Technology Conference** - San Francisco, CA. The conference will cover how Semantic-Based Technologies will be one of the fastest growing areas of Information Technology on the Internet in the next decade. Researchers, academics and practitioners of semantic technologies will be on-hand to answer attendees questions about involvement (and growth) within this sphere. Register via the website, or by phone, 1 (310) 477-4475. Email questions to info@wilshireconferences.com. www.semantic-conference.com
- MARCH 7-10** **Spring 2005 VON Conference & Expo** - San Jose, CA. This event focuses on the convergence of the telecom and Internet industries and the issues effecting the VoIP revolution. Register online, and to get more information call 1 (631) 961-8950 or email von2004@pulver.com. www.pulver.com/von/
- MARCH 11-20** **SXSWWeek 2005** - Austin, TX. The South by Southwest Festival & Conference is comprised of three distinct events: The SXSW Music & Media Conference, The SXSW Interactive Festival, and The SXSW Film Conference & Festival. By day, conference registrants do business in the SXSW Trade Show and partake of a full agenda of informative, provocative panel discussions featuring hundreds of speakers from the international music, film and other media scenes. Register online by February 11 to receive a discount. Questions, please call 1 (512) 467-7979. 2005.sxsw.com
- MARCH 14-17** **Emerging Technology Conference (ETech)** - San Diego, CA. Is there an important technological transformation that you're tracking? If so, you can submit a proposal to lead tutorial and conference sessions at the O'Reilly Emerging Technology Conference (ETech). Deadline for speaker proposals is September 27, 2004. Contact Vee McMillen for more information at 1 (707) 827-7202, vee@oreilly.com. Conference registration begins in November, 2004. For more information contact Gina Blaber, 1 (707) 827-7185, gina@oreilly.com. conferences.oreillynet.com/etcon/ **E**
- MARCH 15-16** **AeA Venture Forum** - Greensboro, GA. The Forum provides private technology companies with a vehicle to access key investment professionals in one location. The program begins with the Technology Industry Golf Event, and

E Events Esther plans to attend.

Lack of a symbol is no indication of lack of merit. The full, current calendar is available on our website, www.release1-0.com. Please contact Kate Tobin (kate@edventure.com) to let us know about other events we should include.

Calendar of High-Tech Events

follows with a day of informative workshops and presentations from companies seeking capital. Register for one or both days via the site, or contact Tina Morais, 1 (408) 987-4234, cristina_morais@aeenet.org. www.aeenet.org/VentureForum

- MARCH 20-22** **PC Forum** - Scottsdale, AZ. For 28 years, PC Forum has been the premier gathering for technology-industry executives, investors, entrepreneurs, thinkers and policymakers. This year's speakers include Marc Andreessen, Mitchell Baker, John Seely Brown, John Thomspon, Jerry Yang, Jonathan Schwartz, Anne Mulcahy, Dawn Lepore...and many more! Register today at www.pcforum2005.com. **E**
- MARCH 21-26** **Doors of Perception 8** - New Delhi, India. This year's theme for the week-long event will be "Infra - Platforms For Social Innovation," and attendees will learn about what infrastructures are needed to enable bottom-up, edge-in social innovation - and how they can be designed. Visit the website to download the registration form, and to keep in contact by subscribing to the mailing list. doors8delhi.doorsofperception.com
- MARCH 22-24** **Cleantech Venture Forum VI** - San Francisco, CA. Also billed as "The Global Conference for Cleantech Venturing" this event aims to be the premiere showcase for venture-grade, emerging clean technology investment opportunities. To register, visit the Cleantech website, or email Lauren Bigelow at lauren@cleantechventure.com. cleantechventure.com
- MARCH 22-24** **The Delphi Proving Ground...** - Boston, MA. ...for Taxonomy & Information Architecture. This is a results-oriented workshop in which participants interact with senior Delphi Information Architecture and Taxonomy faculty. The group workshop setting offers the chance to hear about other teams strategy and design ideas, which are likely to challenge and encourage your team to strive for next-level results as you work toward goals. The three-day workshop repeats in September. Registration can be done via the site, or by calling 1 (800) 335-7440. www.delphigroup.com/events/taxonomy-pg/index.htm
- MARCH 29-30** **World Business Forum** - Los Angeles, CA. The World Business Forum is a symposium featuring nine leaders and thinkers, including Rudy Giuliani, Jack Welch and Anne Mulcahy, speaking about issues and policies of importance to the global business community. Attendees can expect to gain critical insights into the United States' position in the world, management, global financial markets, strategy, the role of change management and other important issues. Register online, or call 1 (866) 711-4476. www.wbfla.com/soundview
- MARCH 30-31** **F2C: Freedom To Connect** - Washington, D.C. F2C is for all who care about - and are affected by - network connectivity, economics, applications and policy. There's a new U.S. Telecom Act in the works, unbundling in Europe, fast fiber in Asia, wireless across Africa and networks being built in cities and villages around the world. "Lead the discussion. Shape the debate. Assert your Freedom to Connect." Register (via the website) before February 28 for a discount. Questions, contact David Isenberg, isen@isen.com. freedom-to-connect.com
-

Visit our new website: More (free-to-read) columns, ideas, essays, features and contributors...featuring Rafe's Radar, a biweekly column by Rafe Needleman. Plus, a new look!

<http://www.release1-0.com>

Release 1.0 Subscription Form

Complete this form and join the other industry executives who regularly rely on *Release 1.0* to stay ahead of the headlines. Or if you wish, you can also subscribe online at www.release1-0.com.

Your annual *Release 1.0* subscription costs **\$795** per year (**\$850** outside the US, Canada and Mexico), and includes both the print and electronic versions of 11 monthly issues; 25% off the cover price when you order from our online archives; a *Release 1.0* binder; the bound transcript of this year's PC Forum (a \$300 value) and an invitation to next year's PC Forum.

NAME _____
TITLE _____ COMPANY _____
ADDRESS _____
CITY _____ STATE _____ ZIP _____ COUNTRY _____
TELEPHONE _____ FAX _____
E-MAIL* _____ URL _____
*personal e-mail address required for electronic access.

☐ My colleagues should read Release 1.0, too!

Send me information about multiple copy subscriptions and electronic site licenses.

☐ Check enclosed

☐ Charge my (circle one): AMERICAN EXPRESS MASTER CARD VISA

CARD NUMBER _____ EXPIRATION DATE _____

NAME AND BILLING ADDRESS _____

SIGNATURE _____

Please fax this form to Brodie Crawford at 1 (212) 924-0240.

Payment must be included with this form. Your satisfaction is guaranteed or your money back.

If you wish to pay by check, please mail this form with payment to: EDventure Holdings, 104 Fifth Avenue, 20th Floor, New York, NY 10011, USA. If you have any questions, please call us at 1 (212) 924-8800; e-mail us@edventure.com; www.release1-0.com.