

Tools for Success II: Don't Search - Filter

Remember why search sucks?

The search engine has gone out to billions and billions of web sites and made an index of the words they contain. *You* enter a word or phrase and the search engine matches those words with words or phrases it has in its index..

That returns way too many, so different engines **use different ways to decide what comes up first**. That is why you see different results in different search engines.

But the foundation that underlies the development of all engines is the same - the idea that **more is better**.

That dates back to the time when there wasn't very much out there.

That has changed. Search engines have not.

The core means of evaluation - the *what* we are paying for - **is STILL get me as many** “relevant” (remember what a weird measure relevance is? Relevant to WHO? When?) documents as you can with as few non-relevant documents.

Word matching.. bad measures... **THAT is why searching sucks** when you are trying to find a specific answer to a specific problem.

How much time do you lose now because of these bad measures?

Do you really have that much extra time to throw away?

Could these moments be the difference between success and failure?

Now, there ARE times I go into a bookstore or a library and just wander to see what is interesting. Search tools as they now exist are fine for that.

However, **when I have a specific problem and I need an answer NOW**, I don't want the card catalogue to tell me what I need is **somewhere** on the 14th floor. **I want to know exactly where it is.**

Here's a new way to get there - **filtering using labels or meta tags.**

First I'll tell you what they are, then what is bad about em and then what is good.

What are tags - a quick 'n simple look:

A card catalogue in a library contains information about the books - the

catalog number, the title, the author, when it was printed, a brief description and **where to find the book. Labels (or meta tags - same thing, bigger name) do the same thing** for a web based object.

An object can be anything that can be accessed on the web - a book, a paper, a copy of a discussion, a picture, a movie... you get the idea.

So what? Well, if the set of labels for an object includes the author's name, you can search for all the things for which that person is the author. For example, online bookstores allow you to search by title, author, ISBN and theme.

If you are skimming.. stop. The next steps are a bit of jump.. and are **VERY IMPORTANT. I want to talk about the different KINDS of tags.**

Object tags:

The tags or labels described above (author, title) are **descriptions of the OBJECT.** Right? Those are pretty straightforward and universal. Anyone looking for a book by Geertz can type in Geertz and the search engine can find that word and return objects that list Geertz as the author. That is fine, unless you want a book by Smith.(22,600,000 returned with a Google of Smith author, 181,112 at Amazon).

Most tags that people talk about are actually this kind of OBJECT tag. They describe the OBJECT - this object was authored by X, it is X pages or bytes, it was created on X date and it is written in X.

The problem with Object tags is that they are pretty general and so won't cut down the number of things returned on a search by a significant number. We need something else.. something that describes the USE to which the object will be put.

Use tag - how is it used?:

What we need are **USE tags** or labels. If I am selling a phone and need to know if I can replace the Ramafrazitiz diode to boost the signal, and if so, how, I need different information than the tech who is repairing the phone or the marketer who wants to make a brochure. If I know that there was a recent discussion of how to answer comparisons with a competing phone and I know that the text of that discussion is in the DOPSS, how do I find it? **Tags or labels that speak to the way the object will be USED are the key.**

However, to be useful, these tags must be

- 1) **unique and**
- 2) **have controlled vocabulary.**

Unique Use Tags:

To have a **simple set of tags** or labels, it is **essential to have**

a set developed for each group - remember the discussion in the last chapter about DOPSS being unique to each group that has a similar information need?

The SIZE of the group doesn't matter.

The distance that separates members doesn't matter.

However, **the kinds of things they need to solve the problems they face must be shared.**

Teachers need lesson plans, repair shops need schematics, salespeople need color options. The kinds of questions and the kind of answers they need will be different.

In order to have a tailored search that returns a useable number of objects (not several million), the Use Tags must be unique.

It's often easier to see an example so below you can see a tag search engine for educators. The bolded captions: What role needs this information, Content area, Grade level, What kind of information do I need, What form or format do I need, and What do I need to do.... are the **Unique Use Tags** for these educators. I'll talk about how to develop these tags later.

Controlled vocabulary for consistency:

In the picture above, the words below each of the tags (What roles,

What role needs this information?	Content Area	Grade Level
<input type="checkbox"/> Teacher <input type="checkbox"/> Administrator <input type="checkbox"/> Planner <input type="checkbox"/> Presenter <input type="checkbox"/> Workshop Leader <input type="checkbox"/> Tech Coordinator	<input type="checkbox"/> Science <input type="checkbox"/> Math <input type="checkbox"/> Technology <input type="checkbox"/> English <input type="checkbox"/> Language Instruction	<input type="checkbox"/> 1-3 <input type="checkbox"/> 4-5 <input type="checkbox"/> 6-8 <input type="checkbox"/> 9-10 <input type="checkbox"/> 11-12 <input type="checkbox"/> Adult
What Kind of Information Do I Need?	What form or Format do I need?	To what use will I put the information? What do I need to do?
<input type="checkbox"/> Facts <input type="checkbox"/> How-to-Guide <input type="checkbox"/> Lesson Plans <input type="checkbox"/> Activity <input type="checkbox"/> Lab <input type="checkbox"/> Research <input type="checkbox"/> Rubric <input type="checkbox"/> Webquest <input type="checkbox"/> Product Comparison <input type="checkbox"/> Links <input type="checkbox"/> Teaching Methods <input type="checkbox"/> Best Practice Example <input type="checkbox"/> Classroom Management <input type="checkbox"/> Evaluation Methods	<input type="checkbox"/> Text Document <input type="checkbox"/> Graphics <input type="checkbox"/> Web Site <input type="checkbox"/> Animation <input type="checkbox"/> Video <input type="checkbox"/> Discussion Board <input type="checkbox"/> Online Tool/Software <input type="checkbox"/> Contact Information/Who is... <input type="checkbox"/> Audio	<input type="checkbox"/> Meeting Standards/Assessment <input type="checkbox"/> Classroom Instruction <input type="checkbox"/> Research <input type="checkbox"/> Presentation/Explaining <input type="checkbox"/> Planning <input type="checkbox"/> Using Technology <input type="checkbox"/> Professional Development - Self <input type="checkbox"/> Adult Workshop

Content area, etc.) are **controlled vocabulary**. The reason it is important to use these controlled (or controlling) words is to make the search (or filtering, a more accurate word) more effective. **It is a way to get everyone to use the same words when adding objects and looking for objects.**

If we don't have controlled vocabulary, everyone uses a different term. I want a lesson plan to teach science in 6th grade. You want the same thing, but you type in "middle school". **In this case, a rose by any other name does NOT smell as sweet because no one can find the thing to give it a sniff.**

The controlled vocabulary is **limitation for consistency**. It makes searching (filtering) faster and more effective.

Remember, we also want to build systems that allow users to teach each other. Essential to that goal is the ability for users to add information into the system. **Controlled vocabulary encourages all the different users to use the same terms when adding something in..fast and effective.**

Yes, controlled vocabulary is limiting. However, this is a necessary limitation in order to **achieve consistency**.

It also makes both entering (adding labels) and finding faster. You just have to click a few check boxes. We develop the vocabulary the same way we develop the unique user tags (more later).

The controlled vocabulary must be allowed to change over time based on users' inputs and needs.

So... you've got objects that live on the web.

You've got some general labels that describe those objects (just like the cards in a "card catalogue").

AND you've got some unique USE tags that are different depending on the group- what kinds of problems they have and what kinds of answers they seek.

You now can do a whole new kind of searching... I like to call it **filtering**.

Filtering— a better way:

When you search, you type in some words. The engine looks at it's index of all the words in all the objects it has indexed (most engines have indexed billions of objects). It matches words and phrases and gives you back a huge pile. Different engines put different things on top, but they all return a big pile.

If you don't type anything in, there is nothing to match, so you get nothing. Simple yes?

Now, filtering works differently.

It STARTS saying, **“Ok, here are all the documents”**. If you don't click any of the check boxes, you get ALL the documents it has -a big pile - **a REALLY big pile!**

If you check something, say **Math** in the **Content Area** tag. Then it does NOT look at the WORDS in the documents. **It looks only at the tags** and **ONLY lets objects through** that have a **MATH** in the **Content Area**.

I think of it like a big vat of stuff. I can open a valve and let everything out, **or I can add a filter** - Math...that will **only let the**

Math “Shaped” things out.

I can add more filters - say a lesson plan filter. So, of the math “shaped” things (which have activities and pictures and discussions AND lesson plans all about math), **only those math shaped things that are ALSO lesson plan shaped can get through BOTH filters.**

If I put in a bunch of filters, I probably won't get anything.

Because the **filters don't look at the words in the objects, only at the tags, the filtration is very accurate.** If I want a science lesson on Stars. I won't get a document with Ringo Starr's plan to sell science lessons.

Tags filter.

Words match.

Tags cut down the pile very quickly and effectively. **When I began using tag searching, it felt like I was using a scalpel when before I had been using a stone ax.**

It's better.

What's bad about em:

The biggest problem with tags are that **they are not automatic**. The big search engines (google, yahoo, A7) all aim toward indexing the entire web. They are always adding documents. **For a tag based filter engine, each document or object must be tagged and added into the system by someone.**

This just won't work if you need to be adding even thousands of things all the time. Many document sets add 100,000 documents a week. A text search engine can do its indexing automatically. It is **the only way to deal with huge piles of information.**

Even with a limited number of tags, it takes a few seconds to tag each object. This won't work for systems that have thousands of new objects being added all the time. It **will never** work for the entire web.

However, for a DOPSS system, where the library of objects is in the thousands and new objects are added in the hundreds, it will work just fine.

The other problem is that **people have gotten used to text search**. It is obvious how it works (**even though it works badly**) and it takes time to get used to using a different method, even if it is better.

Play with this!! **We have built Wherezmy.com to**

demonstrate the concept of a tag based filter/search tool.

There is a **free** version online. Sign on and give it a try.

Join the thousands of people who are discovering the real power of a customized search function.

Once you get the hang of it.. mmmm good!

Whew!!!

A hard bunch of technical! You get a gold star!!

I hope you lived through it.

I wouldn't have made you go through it if it wasn't so important.

We HAVE to get systems where we can file and find information.

It is crucial for success in the information age.

Without it, we will drown in our own swill of information!

We have to spread important ideas and innovations!

It just ain't gonna happen with a Knowledge

Management system that uses awful search engines.

How often have you tried to find something and given up? I do it all the time and I'm pretty good at searching. Mostly, **you need to know what you are looking for to be able to find it**, and if I'm looking for it, it is usually because I bewitched or bewildered... and trying to find something is a sure way to also be bothered.

When we can't find essential information and don't have the time to locate it, we make it up or do without.
This is a sure way to encourage failure.

We need the right tool (information) for the job.

