Organizing Digital Information for Others

How to organize webpages and shared documents
to improve findability

Maish Nichani

maish@pebbleroad.com

About PebbleRoad

We're a user experience design consultancy. We help clients create simple and meaningful websites, intranets and applications.

Find out more about us at http://www.pebbleroad.com

Maish Nichani, Principal of PebbleRoad also published a well-received guidebook for intranet managers titled, **Manager's Guidebook to Intranet Redesign Projects**.

Download the book for free at: http://bit.ly/qhGd9R

Acknowledgements

Table of contents

About this book

If you're part of a web or intranet team then you have to organize information for others—your target users. You have to ensure that they can find the information they need quickly and easily.

Web team or intranet team —> organizing for target users

If you are staff using the intranet for your work, then you have the added responsibility to actively share information **with** others, and organize information **for** others. This is a big shift for you. Till recently, you only organized information for yourself.

Staff —> organizing information for colleagues

The goal of this book is to illustrate some principles and good practice that you could use in organizing information in a way that is meaningful and usable to your target users—the significant others.

Given below is a map of concepts we'll cover in this book. (A higher-resolution version is available on page 51.)
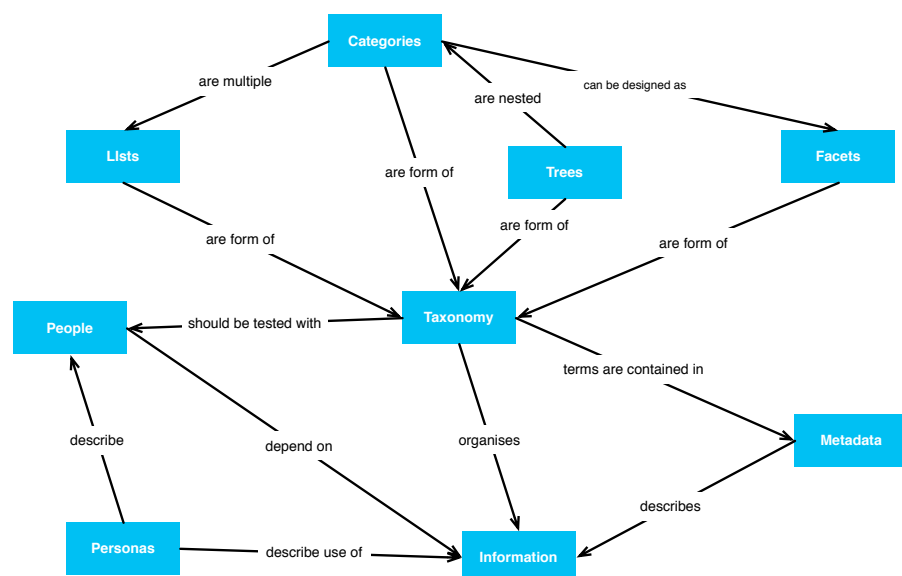

Figure 1. Organizing information—concept map

Why organize information for others?

Digital publishing and social networks have accelerated the amount of information we're exposed to these days. We're publishing more information and are more connected than any other time in history.

The last time such a burst of information took place was after the mechanization of bookmaking by Gutenberg around 1440. We then got **libraries**, **librarians** and **classification systems** to access and manage the growing collections.

The same is happening in today's digital world. Just that the **scales** have changed.

Websites, intranets, shared folders, collaboration spaces	=	the new libraries
Web team, intranet team and more commonly the regular staff (employees)	=	the new librarians
Lists, categories, trees, facets, taxonomies—the stuff you'll learn in this book	=	the classification systems

In 1876, the American librarian Charles Ammi Cutter laid out three rules for a library:

1.   To enable a person to find a book

2.   To show what the library has

3.   To assist in the choice of a book

These rules are largely applicable for digital information as well.

If you're responsible for your company's website or intranet you have an obligation to ensure that your information is found and used. You also have to build in opportunities for serendipity and discovery so that users can stumble upon the wealth of information on your site.

If you're part of a project team managing thousands of documents, then you have to ensure that your team members find the documents easily. However, this notion of organizing for others may be very new to you. For the longest time you've organized information only for yourself—using folders on your desktop. Now, you have to think about others.

In the following chapters you'll learn the various ways of organizing digital information. But first, it's time to learn a little bit about whom you're organizing for.

Understanding your users

How you organize information depends on whom you're organizing for. You would, for example, organize the same information one way for doctors but in another way for patients.

There are many ways to understand the target users. You can observe them, talk to them or learn about them. If the target users are people in your company then you'll be surprised on how much you already know about them.

At the very least, you should be looking for:

- The job they are trying to get done

- The information they need to get the job done

- How they look for this information—the information seeking behavior

Given below is an oversimplified **description of intent,** part of something called a **persona** in design literature.

Joan is an event planner at company ABee. Her immediate job is to plan the up-coming East Asia Forum. She hopes to use last year's event report as a reference. She will look for it on the company's intranet. The keywords she sets off with are: East Asia Forum, annual events, report, 2010.

If people like Joan are target users for your information then such little descriptions of intent can go a long way in shaping how you organize information for them.

An essential tip: **get to know** the people you are organizing information for.

Another tip: don't stop here; **keep testing** your designs with them. You will learn of two tests, **Card sorting** and **Usability testing,** later in the book.

Resources

- **UX Magazine:** Personas: The Foundation of a Great User Experience (http://bit.ly/yQW6FV)

- **UX Booth:** Personas: Putting the Focus Back on the User (http://bit.ly/qqhMS4)

- **Donna Spencer:** Four Modes of Seeking Information and How to Design for Them (http://bit.ly/CmJy)

- **Steve Mulder:** The User Is Always Right: A Practical Guide to Creating and Using Personas for the Web (http://amzn.com/0321434536)
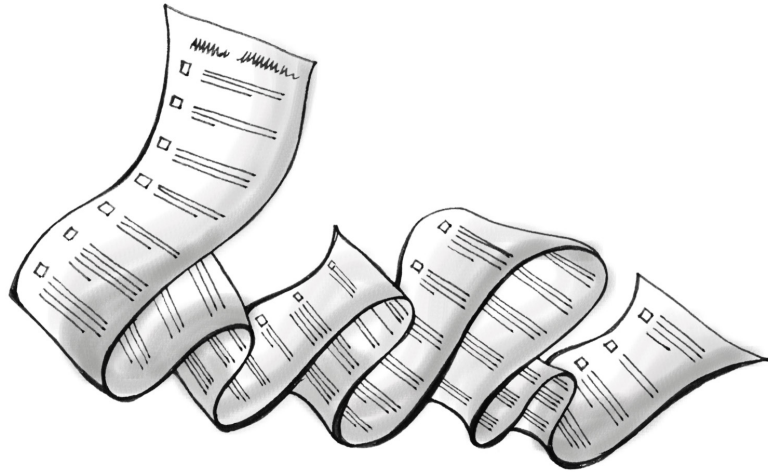
Lists

Lists are a simple and elegant way of organizing low volume information.

We've all made lists at some point or another. We create shopping lists, to-do lists, checklists and reminder lists without even realizing that we're using one of the most natural and efficient information organization devices available to us.

Lists are useful because they group related information together, which can be things or concepts or ideas.

A list of things to buy on the next visit to the grocery store:

- Bread

- Eggs

- Cheese

- Tomatoes

- Carrots

- Milk

- Batteries

A list of webpages on planning an official overseas trip:

- Research and planning

- Budget

- Accommodation

- Air transport

- Land transport

- Local contacts

- Emergency numbers

- FAQs

Although lists are easy and effective when you're the one using it, things can be a little different when you make lists **for others**. This is because others don't have the same context and understanding that you use in creating the list.

Given below is a list of pages on an intranet that show how to purchase products and services:

- ITT

- ITQ

- < 3k

- Reimbursements

- Period

- Outsource

If you've made this list, you know exactly what you're referring to. However, for everybody else, it is a **list of unknowns**!

If you really want to know: ITT is "Invitation to Tender" and ITQ is "Invitation to Quote".

You might argue that everybody in the company would have the background knowledge to understand the list. This can only be possible if nobody leaves or joins the company and those that stay on come to the same conclusion as you. This is a tall order of shared understanding to ask for.

This brings us to the first principle of making lists: **All terms used in a list should pass the common knowledge test**.

Common knowledge

Common knowledge is knowledge that is common among a group of people. The group can be small, like people in a project team, or the group can be large, like everyone in the company.

Here are some examples:

- Driving on the left side of the road is common knowledge in Singapore but not in the US.

- Requiring quotes from three different vendors before awarding a project is common knowledge in government agencies but not in commercial organizations.

- Using the finance system to make large value purchases and requesting the Admin department to make small value purchases could be common knowledge in one company but not in another.

To repeat, the first principle in making lists: **the terms used in a list must be common knowledge to the people using the list**.

A corollary would be: **don't use terms in the list that will make others think**.

The list of pages on making purchases shown previously (Version 1) makes people think. A better version would be something like this (see version 2):

Version 1

Purchases

- ITT
- ITQ
- < 3k
- Reimbursements
- Period
- Outsource

Version 2

Making purchases

- Making purchases—terms you need to know
- Above 70k – Invitation to Tender
- Below 70k – Initiation to Quote
- Below 3k - Small value
- Below 100 dollars (reimbursements)
- Period contract
- Outsourcing

The list on the right has the following elements:

- Organizing principle
- Sequence
- List name, or category
- List terms (title of document or webpage)

In the following sections, we'll go through each of these elements in detail.

Organizing principle

Organizing principle determines how the terms in the list are grouped.

The list on **Making purchases** is grouped according to the **different types of purchases**.

Richard Saul Wurman, the author of Information Anxiety (http://amzn.com/0789724103), proposed that information could be organized in **five different ways**: by

1.   **L**ocation

2.   **A**lphabet (A-Z)

3.   **T**ime

4.   **C**ategory (subject or topic)

5.   **H**ierarchy (continuum or range)

The 5 ways form the acronym LATCH.
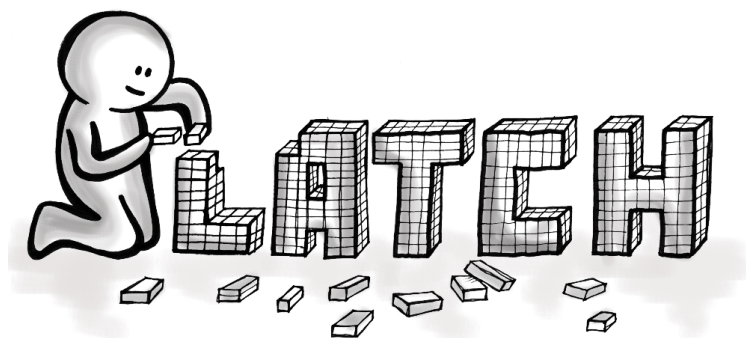
Let's consider an example.

Let's say I want to organize a set of trip reports. Here are the different ways in which I can organize them using LATCH.

By Location

Since trips are made to different locations, I could group trip reports by the different locations they were made to.

Trips made to Asia	Trips made to US & Canada	Trips made to Europe
Trip report	Trip report	Trip report
Trip report	Trip report	Trip report
Trip report	Trip report	Trip report

By Alphabet (A-Z)

Trip reports have a name, so I can use that to group all trip reports by alphabet.

- A1 trip report

- A2 trip report

- B1 trip report

- B2 trip report

- ...

By Time

Trips are made at different times, so I can group them by the year they were made.

Trips made in 2009	Trips made in 2010	Trips made in 2011
Trip report	Trip report	Trip report
Trip report	Trip report	Trip report
Trip report	Trip report	Trip report

By Category (subject or topic)

Trips are made for different reasons, such as for research and study or for conferences, so I can group them by these reasons or categories. Grouping by categories is the most common type of organization.

Trips made for research and study	Trips made for joint visits	Trips made for conferences
Trip report	Trip report	Trip report
Trip report	Trip report	Trip report
Trip report	Trip report	Trip report

By Hierarchy (continuum or range)

When Wurman suggested this way of grouping he was referring to the range of a particular variable, say, the **cost** of the trip. I could, for example, group my trip reports based on how much they cost the company. People auditing the finances of the company will find such a grouping very useful!

Less than $5000	$5000-$10,000	More than $10,000
Trip report	Trip report	Trip report
Trip report	Trip report	Trip report
Trip report	Trip report	Trip report

There you have it: five different ways of organizing trip reports.

But are there really only five ways of organizing information? Not quite. We could add a few more: **by audience** and **by task**.

The Dell website shown below is organized around its audiences.

Figure 2. The Dell website is organized around its key audiences

The help section in *Xero*, an online accounting application, is organized by task.

```
⊟ Invoicing in Xero
    ⊟ Accounts Receivable (customers)
        📄 Add or edit an invoice
        📄 Create or edit a repeating invoice
        📄 Invoice branding: customize invoices
```

Figure 3. The help section in Xero (http://xero.com) is organized around tasks

Organization and meaning

Going back to trip reports, did you notice that each way of organizing gave a specific view and meaning to trip reports? Organizing by location gave a view and meaning that organizing by budgets did not.

So, which way of organizing is better? This really depends on why you are organizing and for whom. If you were organizing trip reports for a finance officer in charge of budgets, then organizing by hierarchy (by cost range) may work.

Tony Pritchard describes such a search for meaning in organizing the names in The Vietnam War Memorial in Washington DC. See http://bit.ly/sfXNku.

Organizing the names alphabetically would depersonalize the lives of the soldiers. Organizing by rank (category) would have the same outcome. The authorities therefore decided to organize the names of the soldiers based on those they died with. As Pritchard goes to say, "Any other organization would have altered the meaning and form of the memorial".

Let's move on to **Sequence.**

Sequence

You must have guessed this. Sequence is the order in which the terms appear in the list.
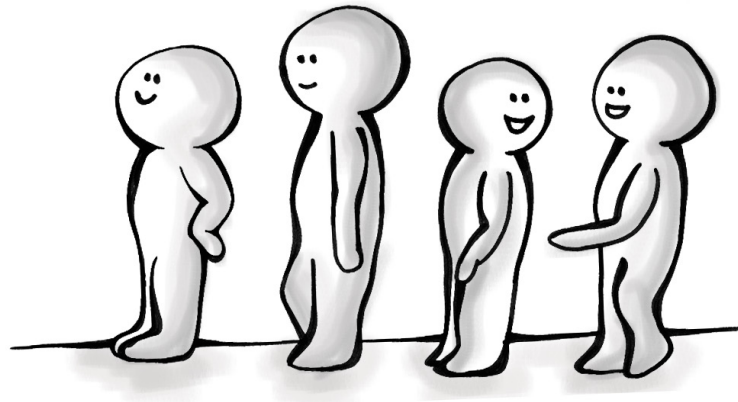
Here is a simple two-step technique you can use to determine sequence.

1. Check if there is an inherent logical sequence. For example:

    a. Latest on top (by date)

    b. First to last

    c. Frequently to occasionally

    d. Simple to complex

    e. Known to unknown

2. If you can't find an inherent sequence, go for alphabetical (A-Z) sequencing

If, however, you think there is a structure and you can't pick a winner, do a card-sorting test (see the chapter on Card Sorting).

List name or category

After you've considered the organizing principle and the sequence of terms, it's time to give the list a good name. Without a name there is nothing holding the list together. A list of documents does not mean much unless you give it a name such as **trip reports**.

A list name is also referred to as a **category**. The dictionary definition of a category goes something like this: "a class or group of things, people, etc., possessing some quality or qualities in common".

Let's consider an example.

In the list below, **Broccoli** belongs to the **Vegetables** category.

Vegetables

- Broccoli

- Carrots

- Cauliflower

Each term in the list has the same relationship to the category—they are all veggies. If I add 'banana' to the list then the relationship will break—it's a fruit and not a vegetable.

Now, this is not a revelation—we're exposed to this kind of stuff as kids. Remember the activity called 'Pick the odd one out'?

Let's take a look at the Vegetables category again. If I change the category to **A healthy diet**, then banana can be on the list. The relationship is now correct and intuitive.

A healthy diet

- Bananas

- Broccoli

- Carrots

- Fish

- …

Some categories are so common that we know exactly what they contain such as the Vegetables or Pop Music category. But will people know what exactly goes under the **trip reports** category?

This gives us another principle of categories: **like the terms in a list, categories too are subject to common knowledge**. The category Trip Reports may seem out of place if staff seldom go on trips or don't write trip reports.

Now, let's assume that category trip reports passes the common knowledge test. But is the name the best we can come up with? Or are there better ways of naming categories?

As with everything else, there are guidelines and principles that can help us along the way.

Good names:

- clearly describe the terms in the list

- use simple everyday words

- use verbs where possible

- are concise

Web writing expert Ginny Redish (http://1.usa.gov/4ab6B5) lists three types of names:

Type of name	What it is	Example	Scope
Question	A name in the form of a question	How do I plan an overseas trip?	Narrow
Statement	A name that uses a noun and a verb	Planning an overseas trip	Broad
Topic	A name that is a word or short phrase	Overseas trips	Broader

The Question and Statement types work well on websites and intranets. This does not mean that you should avoid using Topic type. On the contrary, if the terms in the list are very familiar to your group then topical names will work quite well.

Here are some topic names used for listing staff benefits:

- Medical

- Dental

- Insurance

What does all this mean? Before we come up with categories for others to use, we need to be mindful of the common knowledge effect and take this into account when creating categories.

When in doubt do a card-sorting test (see chapter on Card Sorting).

List terms

On the web, each term in a list is usually a link to a webpage or document. The term therefore takes the name of the webpage or document it points to. If the category makes sense but not the terms, then this will confuse people and defeat the purpose of the list. You now see why it is important to have a good naming practice for webpages and documents.

Given below are three ways of naming webpages and documents. Notice that the same principles of naming categories are used.

Version 1 (Question)	Version 2(Statement)	Version 3 (topic)
Excellent service award	**Excellent service award**	**Excellent service award**
• What is it?	• About the award	• Award
• Who is eligible?	• Check if you are eligible	• Eligibility
• How to apply?	• Submit your application	• Application
• Who has won this before?	• View gallery of past winners	• Past winners

Did you notice that terms under a category are of the same type?

The terms do not change from a Question type to a Topic type. This is called **parallelism**. When naming terms, try to keep to one type. Using different types just confuses people.

Here is the same list using different name types.

- Excellent service award

- What is it?

- Checking if you're eligible

- Application

- Who has won this before?

Most likely the list slowed you down a bit. It's because you had to shift between different name types.

Resources

- **Ginny Redish:** Letting Go of the Words (http://bit.ly/xtg6yY)

- **UX Booth:** Classification schemes and when to use them (http://bit.ly/aUcQPx)

- **Richard Saul Wurman:** Business of Understanding (http://bit.ly/u29Wpw)

- **Nathan Shedroff:** Organizing things (http://bit.ly/4XXyRO)

- **David Weinberger:** Everything is Miscellaneous (http://www.everythingismiscellaneous.com/)

Trees

The previous chapter was all about single categories. But categories can be nested as well—**categories under categories**. And this nesting can go deep. Just look at your PC's file manager for examples of deeply nested categories.

These nested categories are called **trees**. We can also refer to them as hierarchies, but hierarchies imply an inheritance property.

In the example given below, a Square inherits the qualities of the broader terms. A Square is-a Quadrilateral, is-a Polygon and is-a Shape.

- Shape
    - ↪ Polygon
        - ↪ Quadrilateral
            - ↪ Square

In the example below however, Privates do not inherit the qualities of the higher chain of command. A Private is not a Sergeant, for example. This arrangement is called a **tree**. Trees offer a more lenient approach to nested categories.

- Generals
  - ↪ Colonels
    - ↪ Captains
      - ↪ Lieutenants
        - ↪ Sergeants
          - ↪ Privates

Shown below is an example of a tree used in intranets.

- Some company
  - ↪ Corporate affairs
    - ↪ Human resources
    - ↪ Recruitment
    - ↪ Admin
    - ↪ Welfare
    - ↪ Training
    - ↪ ...
    - ↪ Finance
    - ↪ Computers and technology
    - ↪ ...
  - ↪ Products
  - ↪ Sales
  - ↪ Marketing
  - ↪ Research
  - ↪ Engineering
  - ↪ ...

Here is a question: **Why do we have trees? Why can't we live with multiple, independent categories?**

If you had only a few items to arrange, you'd simply use a list. If you had more than a few, say 15-20 items on the list, then you'd categorize them into logical groups. If you had around 200-500 items on the list, you'd definitely categorize them but you may start finding a need to nest them as well, thereby building a tree. Maintaining multiple categories at this point just makes the arrangement unwieldy. Just take a look at the folder structure on your computer—a classic example of a tree structure.

The main purpose of trees is to organize high volume information so that we can understand and use them more efficiently. It is the same reason we have corporate structure (a tree)—to command and manage the company and drive it to success.

A folder structure on your computer is still manageable as you're the one creating and nesting the folders. But what happens when you want to share your folder structure with others?

Your folder structure may not go down well with others, unless and only if, the principle behind the arrangement is common knowledge. The only way to know for sure is to test it out.

Did you notice how many times we're resorting to **card sorting?**

Shape of a tree

A common question on trees has to do with its shape: **Should it be narrow-and-deep or broad-and-shallow**?
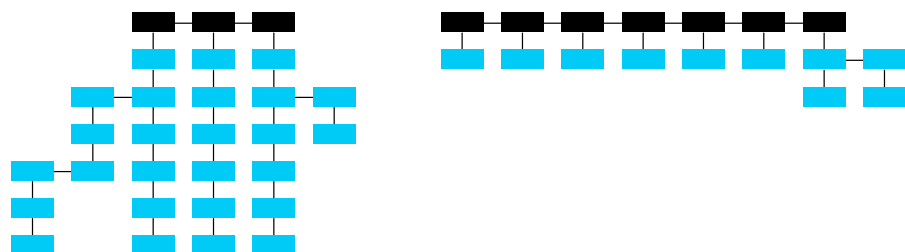

Figure 4. Narrow-and-deep (left) and Broad-and-shallow

The shape of a tree affects the **predictability** of finding information on it.

You may say that a broad-and-shallow tree shows more of it at first, so it is more predictable. However what happens if the categories are similar or ambiguous in meaning? Users will have a hard time figuring out which categories to go with.

Or you may say that a narrow-and-deep tree offers users fewer entry points so it is more predictable. Not quite right here too. What if the deeper levels are used more frequently? Users will need to make many decisions before getting to the right level.

So, back to the original question: narrow-and-deep or broad-and-shallow? The answer: it depends!

But here's a tip: **start with the broad-and-shallow shape first**. Just be careful with overlapping or confusing categories. Broad-and-shallow improves the visual scanning of information, something that people do often when browsing.

Problem with trees

Let's take an example to understand the problem with trees. Let's say you've created a tree for organizing **trip reports**.

Staff go on trips and when they come back and write a report on what they did, what they learned, whom they met and so on. This is called a trip report. The trip report becomes the diary for the trip and an important source of knowledge that can benefit others.

The purpose of creating the tree structure is to help:

- Staff submit their trip reports

- Staff find previous trip reports

If you were in charge of creating the tree structure, how would you create it?

If you want to make it easy for staff submitting the report, you might organize the folder by year:

- 2008

- 2009

- 2010

Your organizing principle could be: **by the year the trip was made**.

So far, it looks like a nice organizing principle. But what happens when a staff comes along looking for a report on a trip made to Australia?

Well, she better know the year the trip was made! Otherwise, she'll be spending time (and company money) opening every category and looking for the report on the Australia trip.

Organize by 'country' and by 'year' then, you say. That should do it. Let's see.

- Australia

  ↳ 2010

  ↳ 2000

  ↳ 2008

  ↳ ...

Cambodia

  ↳ 2010

  ↳ 2009

  ↳ 2008

  ↳ ...

Chile

  ↳ 2010

  ↳ ...

This might serve both staff submitting the report and staff seeking the report. However, there is still one problem. You knew this was coming!

Trips are rarely made to chalk up visits to different countries. They are made to serve work-related goals. They may be trips made for **academic conferences**, **research collaboration** or **joint exercises**. Now, if there are only a few such trips made in a year, then we're still fine with the tree structure by country.

But if many such trips are made, then you may have the possibility of staff coming along looking for 'joint-exercises in 2008'. If you have many such queries, then you're in trouble. Staff will have to do the unproductive work of opening and checking every document under 2008 to find reports that were **joint-exercises**.

You only wish you could have information magically reorganize itself into different tree structures on demand—one by year, another by country and yet another by purpose. Only then you could satisfy all types of queries (at least a large majority of them).

Actually, you're in luck. There is a way to create this on-demand organization of information. It is called organizing by **facets**.

Resources

- **UX Booth:** Classification schemes and when to use them (http://bit.ly/aUcQPx)

- **PebbleRoad:** Designing Site Structures for Intranets and Websites (http://bit.ly/upIe2o)

- **Rosenfeld Media:** Card sorting: designing useful categories (http://bit.ly/eAzQN)
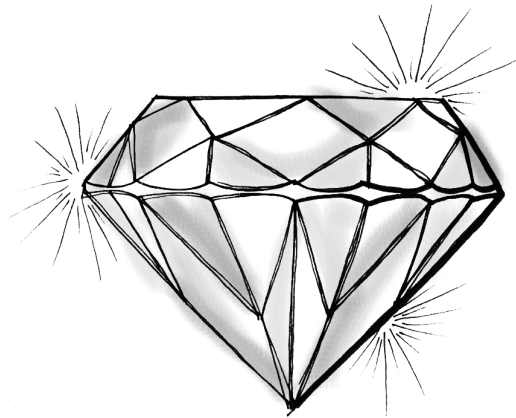
Facets

A facet is a particular view to information. Information can have many facets.

Let's get back to our example on trip reports. You've already picked out many facets or views to the reports. These are:

- by year

- by country

- by purpose

Note that facets have values as shown:

- By year

    ↳ 2010

    ↳ 2009

    ↳ 2008

    ↳ 2007

- By country

    ↪ Argentina

    ↪ Australia

    ↪ Brazil

    ↪ South Africa

- By purpose

    ↪ Academic conference

    ↪ Research collaboration

    ↪ Joint-exercise

Every trip report can now be assigned values from the list above. We then get a table like the one shown below:

Document	By year	By country	By purpose
trip-report1.doc	2010	South Africa	Research collaboration
trip-report2.doc	2010	Australia	Research collaboration
trip-report3.doc	2009	Brazil	Academic conference
...	...	...	...

Here are a couple of points to note about the table:

- It is very different from a folder structure—it is one big table

- Unlike the folder structure, this table does not really include the physical document; it just gives a reference to the document, which might reside on a document server

- Every document is assigned values from the relevant facets

Facets offer significant advantages over trees:

- You can **start with any view**—by year, country or purpose. With trees, you don't have a choice—you're stuck with the view chosen by the person who created the tree.

- You can **navigate faster** by choosing a combination of views— show all trip reports made to Australia in 2008 that were for academic conferences.

When you select facets, you are filtering or narrowing the collection to show only the documents you're interested in. Let's take a look at a few examples to see how this works.

Filtering by facets

Explore how these websites offer facets to help users quickly filter the collection and find what they want. (Open a browser window, visit these sites and play with the facets).

Figure 5. Wine.com (www.wine.com)

Figure 6. Artist Rising (www.artistrising.com)

Figure 7. Kayak (www.kayak.com)

Here are some points to note:

- For Kayak and Artist Rising, the filters only appear after a search is made.

- For Wine.com, however, the filters appear as a browsing option.

- You can start with any facet you want.

- There are two popular ways to place the facet navigation on the webpage: left or top.

- There is a feedback bar that shows the current selection. There is an option to clear the selection as well.

- The big challenge is when facets have nested categories (see screenshot below).

Figure 8. Navigating facets with nested categories

Choosing facets

How do you pick the right facets for your information? By understanding how others would like to find your information.

Experts in the field of information science have identified several facets that seem to be applicable to many types of information across many disciplines. These are listed below. You can use these as a reference or starting point.

General facets to information:

- by the thing, the kind or type of thing, or its properties

- by time

- by location

- by author

- by activity, task or operation

The facets for trip reports (the thing) match up as follows:

- year (by time)

- country (by location)

- purpose (by kind of thing)

There is an important point to note about facets: **they must be mutually exclusive**. You cannot, for example, have a facet for Regions (North America, Asia, etc.) and another facet for Countries.

- Countries
    - ↳ Australia
    - ↳ Singapore
    - ↳ …
- Regions
    - ↳ North America
    - ↳ Asia Pacific
    - ↳ …

You might think of combining the two into a single facet, say **Locations**. Even then, you cannot have terms that overlap in meaning. In the example below, how would you classify a trip made to Singapore? Will it be under Singapore or under Asia?

- Locations
    - ↪ North America
    - ↪ Europe
    - ↪ Asia
    - ↪ Singapore
    - ↪ Hong Kong
    - ↪ Australia and Oceania
    - ↪ Middle East

This does bring up an interesting scenario: What if you **nest** the terms?

The following is valid:

- North America
- Europe
- Asia
    - ↪ Singapore
    - ↪ Hong Kong
- Australia and Oceania
- Middle East

In other words, facet terms can be lists or they can be categories. Keep this in mind and you'll be fine!

When to use facets? When to use trees?

Trees are used when a single dimension or view of the organization is enough. For example, consider this structure:

- Staff handbook
    - ↳ Leave
        - ↳ Annual leave
        - ↳ Sick leave
    - ↳ Benefits
        - ↳ Medical benefits
        - ↳ Insurance plans

Now, you may want to **control this structure**. You want all the types of **Leave** to be in a particular order and you want **Insurance** plans to come after **Medical benefits**. You may also want all staff to see this single view. In such a case, it may be better to use a tree.

If on the other hand, let's say that you have 5,000 documents and a deliberate structure between the documents does not matter, then we can help staff find the documents using appropriate facets.

This brings us to the principle: **if you want to present staff with a single, deliberate structure then use a tree. If you want staff to access the information from multiple views then use facets**.

Resources

- **PebbleRoad:** Designing collections for the web (http://bit.ly/cfQRpv)

- **A List Apart:** Design Patterns: Faceted Navigation (http://bit.ly/dmn20m)

- **UX Matters:** Best Practices for Designing Faceted Search Filters (http://bit.ly/10ibD2)

- **Step Two Designs:** Using metadata and taxonomies to classify and filter intranet content (http://bit.ly/ezxlRn)

Metadata

To understand metadata just ask yourself this question the next time you go shopping: **can I get all the information I need from this product I'm holding in my hand to make a well-informed purchase decision**?

If you got all the information you need, then most probably you got the information from the label on the product. The label on the product shows the brand, material, price and any discounts applicable. The information on the label is **metadata**.

It is the same with digital information. If you're looking at a document and you can get the **Title**, the **Author**, a **Description** and the **Date** it was published, then you're looking at the document's metadata.

There are two things you need to know about metadata:

- Metadata is not usually part of the information contained in the document (though David Weinberger argues in his book, *Everything is Miscellaneous*, that everything in the document is metadata)

- Metadata contains rich information that **identifies** the document (e.g. Title, Author, Publish Date etc.)

Here's a question: What happens if 10 documents have an **Author** metadata and 10 other documents have a **Creator** metadata to refer to the same thing—the person who authored the document? How will you find this document?

To give good results, the search algorithm has to know that **Author = Creator**.

Now, just imagine the overheads that the engine must bear with if there are hundreds of such lookups. It makes the use of metadata seem inefficient.

A simple solution exists: **be consistent with metadata**. Ensure that you use the same metadata to identify or describe the same type of information.

Question: Is it possible to get consistent metadata values that can be applied to all types of information in the company?

Yes. But there's more to it.

First, not all information requires all metadata. Draft or personal documents, for example, will need little metadata, if any at all.

Second, for information that needs metadata, we can start with a generic metadata set. You've already seen glimpses of it in the previous paragraphs. Title, Author, Description, Publish Date, etc., are a generic set of metadata that can be applied to any type of information.

Such metadata is so generic that there is a standard known as **Dublin Core** that describes 15 metadata elements that can be used for all types of information. The fifteen elements are Contributor, Coverage, Creator, Date, Description, Format, Identifier, Language, Publisher, Relation, Rights, Source, Subject, Title, and Type. See http://dublincore.org/documents/dces/.

After the generic set, we'll need another set **specific** to the needs of the company. It covers metadata such as Approvals, Workflow, Security and Preservation. This requires a bit of research and study. An external consultant is usually sought for this kind of work.

So there you have it: metadata = **generic set + specific set.**

Assigning metadata

Now that the metadata set is specified for the company, how will staff assign metadata to the information they create? The assigning is usually done using the Content Management System (CMS) you use for the website or intranet.

Assigning metadata can be frustrating if not planned properly. When staff complain that "the system puts up this screen and expects me to fill out 20 fields", they are really complaining about the metadata assignment strategy.

There are many ways to make the assignment simple and usable. We'll cover just three here:

System assigned

The CMS can automatically assign some values. For example, Author, Publish Date and File type can automatically be applied by the system.

Smart defaults

Smart defaults try to predict the context of the user and present values that it thinks will be relevant. For example, the CMS might surface my commonly used metadata values or only those metadata applicable to the user's department or job function.

Smart templates

Templates can have pre-assigned metadata. Authors only need to fill out the remaining metadata. This can speed up the authoring process.

Tag bundles

Tag bundles takes the smart defaults forward in a creative way. If you keep assigning the same values to information you create then instead of repeating your work, you assign the values to a bundle and then just apply the bundle to the information. See http://bit.ly/uk4XOH.

Metadata and controlled vocabularies

Let's say you ask staff to fill out the **Purpose** of the trip when submitting trip reports. What will they fill out? Some might put in '**business**' while others might put in '**JE**' for joint-exercises. You see the problem. There will be no consistency in what staff enter as metadata. How can we then get consistent metadata values?

By controlling what staff can choose.

A **controlled vocabulary** is a controlled list of terms. Let's say you give staff only three options to fill out the **Purpose** of the trip. These options could be:

1. Academic conferences

2. Joint exercises

3. Research collaboration

Staff can only select from these three options. The values are **controlled**.

But what if the controlled vocabulary is itself confusing and ambiguous? It is a serious problem, which points again to a lack of a metadata strategy.

A controlled vocabulary can do more than just control the selection of terms for authors. They can also enable some backend magic during search and retrieval.

Let's say a staff searches for the term "**JE**", the informal use of the term "Joint Exercises". Trip reports, being formal documents, do not contain the term "JE". They are assigned the metadata term "Joint Exercises". The result is that the search engines will not find the documents, even though they exist.

This is where an expanded controlled vocabulary can shine through. Instead of just listing controlled terms, it can also include the following:

JE USE Joint Exercises

The statement tells the search engine to use the term "Joint Exercises" when it comes across the term "JE". Problem solved.

You can imagine more situations where this simple expansion could help such as homonyms, misspellings and abbreviations.

But wait. There's more!

A controlled vocabulary can be expanded in more ways than just USE:

NT = narrower term (Marketing NT Corporate Marketing)

BT = broader term (Invitation to Quote BT Corporate Procurement)

RT = related term (Trip report RT Country Report)

We are putting a structure around the linear list of terms in the controlled vocabulary, thereby creating a hierarchy. This kind of controlled vocabulary is called a **thesaurus**.

Resources

- **Boxes and Arrows:** What is a controlled vocabulary? (http://bit.ly/10Do9v)

- **Getty Research:** Introduction to Metadata (http://bit.ly/umbpRS)

- **Getty Research:** Introduction to Controlled Vocabularies (http://bit.ly/fYjFQF)

- **Niso Press:** Understanding Metadata (http://bit.ly/tVmuuc)
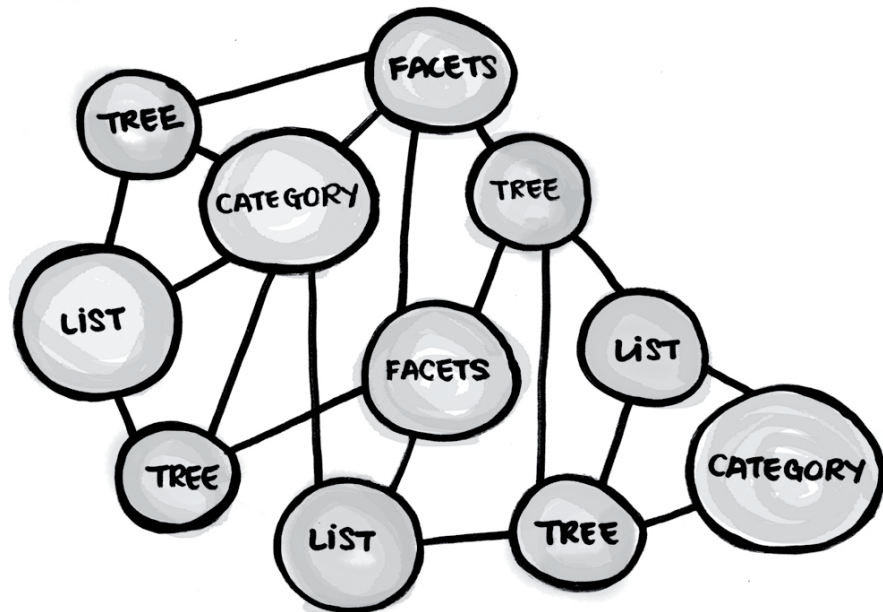
Taxonomy

A **taxonomy** is a structured way of organizing information. It comes in different shapes and sizes. And it is not always hierarchical.

Yes, you read that right—**taxonomy is not always hierarchical.**

In fact, the different ways of organizing information that you've seen in this book—lists, categories, trees, facets—are **shapes** a taxonomy can take.

Patrick Lambe has written an illustrative post titled "*What shape is a taxonomy?*" (http://bit.ly/ymWNaa) where he goes on to describe why treating a taxonomy as being only hierarchical may be misleading.

If a taxonomy is a structured way of organizing information, and if it can take on many shapes, then in a company we're really looking at a **system of taxonomies** all working together. This is usually called a **corporate taxonomy**.

So, a corporate taxonomy is not one gigantic structure for organizing all corporate information. It is a system of multiple taxonomies.

The glue that connects the system is metadata. Common metadata connects taxonomies.

Think about these two points the next time you're organizing information:

- The shape of your taxonomy will help others easily look for your document

- The metadata you apply will surface your document in places (other taxonomies) and in situations that you would have never guessed, thereby enabling knowledge discovery

Taxonomy and metadata

Taxonomy is how information is structured while metadata contains terms that describe the information.

Taxonomy terms usually appear in metadata in the form of subject terms or category terms. This is how a taxonomy can link up with other taxonomies.

- Title

- Author

- Date

- Category <-- From taxonomy

Note that you can have more than one **Category** metadata field. If you are using several facets, you will have a category metadata field for each facet.

Taxonomy and tags

Tagging is about assigning personal metadata to information. It comes from the practice at popular websites such as delicious and Flickr.
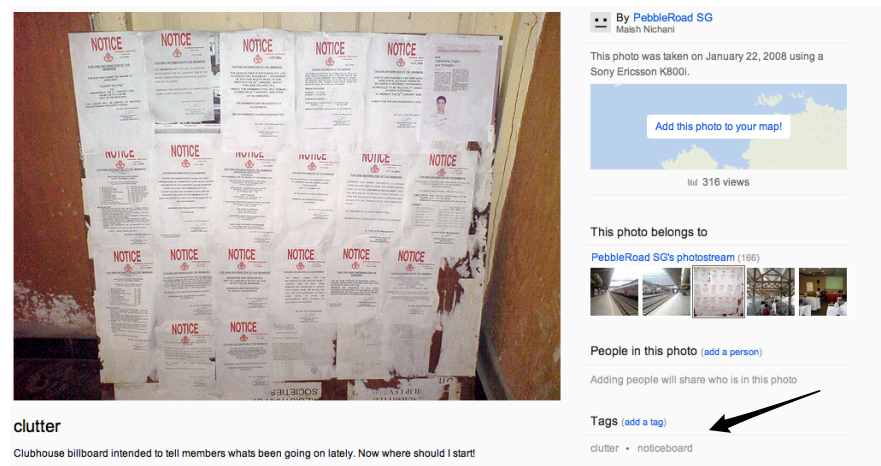

Figure 9. Tags in Flickr

The tags that I used for the picture are "clutter" and "noticeboard". They are not controlled terms. I made them up. I chose the terms because I thought that the picture might be found by using the two keywords. I could be right or I could be totally wrong. I could be right if there are other people who think about it as a "noticeboard". I could be wrong if many people think about it as a "billboard". Now, I can simply add "billboard" to the list of tags. But then, there could be more such terms out there.

The effectiveness of tags therefore goes back to the level of common knowledge. If more people understand the tag terms the more effective they will be.

Tags can be captured and used as metadata. The difference is that they are not controlled.

Tags can be used in two ways:

1. Free tags

2. Semi-controlled tags

Free tags are like the tags used in Flickr. No attempt is made to find or check if the tag is correct or relevant. With free tags it will be difficult to link up different taxonomies.

Semi-controlled tags have some sort of check before assignment, which goes something like this:

- When someone uses a new tag, it is remembered (e.g. "noticeboard")

- The next time someone starts typing "noti…" the system suggests "noticeboard" as a term, thereby promoting consistency

This can get pretty interesting with the following:

- Suggestion of controlled terms

- Auto-replace, which will correct "Phillippines" to "Philippines" automatically

- Suggest synonyms (suggests "noticeboard" when "billboard" is used)

Semi-controlled tags attempt to exert some control, but not complete control. Tags present an interesting opportunity and challenge and must be handled with care.

Resources

- **Green Chameleon:** Defining Taxonomy (http://bit.ly/bsNjFa)

- **Green Chameleon:** What shape is a taxonomy? (http://bit.ly/ymWNaa)

- **Patrick Lambe:** Organizing Knowledge (http://www.organizingknowledge.com/)

- **Barbara Kwansik:** The Role of Classification in Knowledge Representation and Discovery. (http://bit.ly/wdpuDk)

- **Heather Hedden:** The Accidental Taxonomist (http://amzn.com/1573873977)

- **Gene Smith:** Tagging: People-powered Metadata for the Social Web (http://amzn.com/0321529170)

Testing

Organizing information for others is a big responsibility. Your design can influence other people's productivity and understanding when they visit your page, website or intranet. The key to becoming good at organizing information is to be reflective and **challenge the decisions you make**. Do I have the right organizing principle in place? Have I selected the right shape of the taxonomy? Is it usable? What's stopping the information from being found? You can get answers to these questions by constantly testing your ideas and designs with target users.

Two techniques are used often to run such tests: **card sorting** and **usability testing**.

Card sorting

Card sorting is a simple exercise you can use to come up with categories or test if the categories are working.

Since trees and facets also use categories, card sorting also helps in testing trees and facets as well.

The technique goes something like this:

- Gather all the terms you want to test.

- Write down each term on a yellow colored index card. Again, you can use any color. Use one card per term.

- Place all the cards on the table.

- Invite participants. Make them feel comfortable.

- Ask the participants to group the cards that they think belong together.

- Ask them to name their groups.

- Analyze the results and compare it with any initial structure you had in mind.

The type of card sorting mentioned above is called **open card sorting**. It is open in the sense that there are no pre-defined categories.

Closed card sorting on the other hand is when we do the exercise with pre-defined categories. Here's how it might go after putting

the yellow cards on the table:

- Jot down the name of each category you want to test on a red color index card. You can use any color. Use one card per category.

- Ask the participants to match the yellow cards to the categories they think the card belongs to.

- Analyze the results and check if the categories are intuitive.

As you would have guessed, open card sorts are great for coming up with shared categories because you're working the common knowledge of the participants.

If participants are not at the same location, then you can use online tools such as **TreeJack** (http://www.gettreejack.com/) to do remote card sorts. These tools also do the math for you and present the findings visually. Talk about making things simple!

Resources

- **Boxes and Arrows:** Card sorting: a definitive guide (http://bit.ly/16rTpL)

- **IQ content:** How to: Card Sorting (http://bit.ly/9KQtzO)

- **Rosenfeld Media:** Card sorting: designing useful categories (http://bit.ly/eAzQN)

- **UX Booth:** Classification schemes and when to use them (http://bit.ly/aUcQPx)

Usability testing

Even if you know the different ways of organizing information for others, your design may not be usable. Sad but often true.

Usability is about making it easy for users to use your design. The design community has over the years given us some good guidelines that we can follow. Some these are shown on the next page.

However, even after using the guidelines, you may find that users are confused with your design. But what could be causing the confusion? To find this out, you need to put your design to a test. Such a test is called a **usability test.**

A usability test goes something like this:

- Know why you are testing (create the hypothesis)

- Write down scenarios that will give data to test your hypothesis

- Create the page or pages you want to test

- Recruit the right users to take the test

- Run the test and record the session using a video recording software such as Silverback (http://silverbackapp.com/)

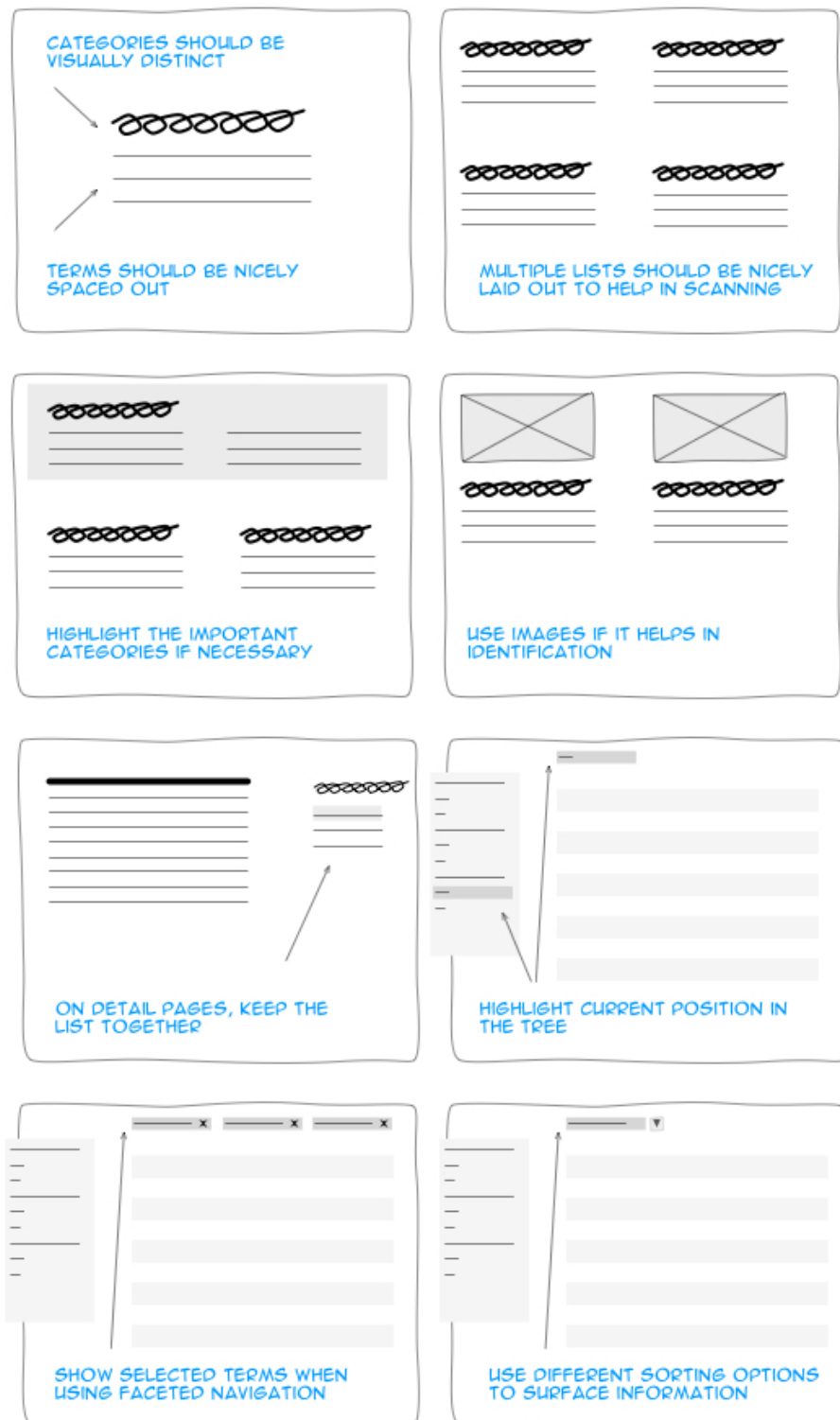- Thank your participants and start analyzing the results

Figure 10. Making information organization usable

When considering usability issues, do not ignore the authoring and publishing interfaces. Many times usability issues that end users face (front stage issues) are a result of the problems that the authors face (back stage issues).

Resources

- **Jakob Nielsen:** Alertbox (www.useit.com/alertbox/)

- **Steve Krug:** Don't Make Me Think
  (http://amzn.com/0321344758)

- **Usability.gov:** (http://usability.gov/)

- **Ronnie Lipton:** The Practical Guide to Information Design
  (http://amzn.com/047166295X)

- **Jeffrey Rubin:** Handbook of Usability Testing:
  How to Plan, Design, and Conduct, Effective Tests
  (http://amzn.com/0470185481)

Conclusion

From lists to trees and from metadata to taxonomy, we've covered a lot of ground in 50 pages. Although we took an introductory approach to these subjects, I hope it gives you the confidence to have a conversation with your team and colleagues on the importance of organizing information for others. These conversations in turn can help make our pages, websites and intranets more effective and the people using it more happy!

So, go ahead and spread the message and let the conversations begin!

Short link to the book:

http://bit.ly/yEyfFZ

E-mail link   Tweet link

Before we close here are two takeaways to keep by your side:

1. A map of the concepts covered in this book

2. A checklist to jog your memory when organizing information
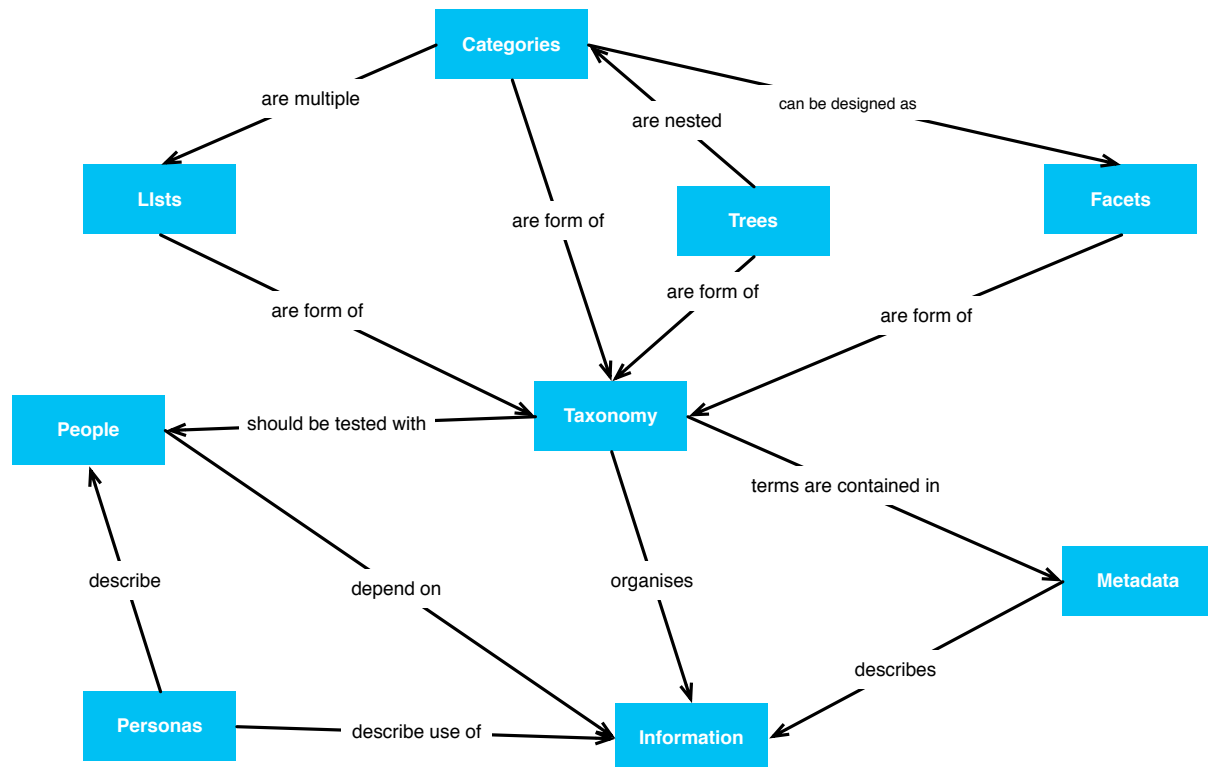
Happy organizing!

Concept map

Figure 11.   Organizing information—concept map

Checklist

For lists and categories…	
☐	Goals and needs of the target users are clearly defined
☐	Organizing principle is intuitive
☐	Terms and categories pass the common knowledge test
☐	Terms are arranged in the right sequence
☐	Common layout issues are taken care of (spacing, layout, legibility, contrast)
For trees, add these…	
☐	Shape of the tree is well thought-through (narrow-and-deep or broad-and-shallow)
☐	Structure is tested with target users
☐	Interface is usable and intuitive
☐	Useful sorting options are available (especially for large collections)
For facets…	
☐	Organizing principle is intuitive
☐	Facet categories are mutually exclusive
☐	Facet terms are mutually exclusive
☐	Facet categories and terms pass the common knowledge test
☐	The structure is tested with target users
☐	Interface is usable and intuitive
☐	Useful sorting options are available (especially for large collections)
For metadata	
☐	Metadata is consistent (based on a framework)
☐	Values are controlled where possible (controlled vocabularies)
☐	Auto-assignment is supported where possible
☐	Authors are trained on assigning metadata
☐	It is easy for authors assign metadata