

# Netsorting

## User centered Information Archi

Stefano Bussolon, Fabio Del Missier

[stefano@bussolon.it](mailto:stefano@bussolon.it)

Università di Trento, Dipartimento di Scienze della C  
Formazione, via Matteo del Ben 5, I-38068 Roveret

# Summary

- Information Architecture
- Structure
- Labeling
- Free listing
- Card sorting
- Netsorting
- Conclusions

## Cognitive artifacts that

- Visual appearance and impact.
- Utility: the ability of the web site to do fu  
what it is supposed to do.
- Usability: how effectively a person can use

## Web sites main purposes

Web sites often serve two primary purposes: to provide information and to perform a task

The utility of a web site is given by its capability to provide the information required by the users or to allow them to perform the tasks they want to perform.

The usability of a web site is provided by its interface, i.e. the way the information or the tasks are structured and labelled.

Web usability concerns with two linked but distinct dimensions:

- the web interface;
- the information architecture.

## User Centered IA

An user centered information architecture (UCIA) can increase both the utility and the usability of an information system like a web site.

UCIA can improve the utility of a site because user centered methods elicit the needs, expectancies and intentions of the participants;

UCIA should improve the usability of a site because user centered methods elicit the users' lexicon and their mental models about how they expect the information is structured.

## Definitions of IA (1)

IA: some definitions

Andrew Dillon: "IA is the term used to describe designing, implementing and evaluating information systems that are humanly and socially acceptable to the stakeholders."

"simply a set of aids that match user needs with resources" (Davenport, 1997)

"a structure or map of information which allows their personal paths to knowledge" (Wurman, 1989)

## Definitions of IA (2)

“Information architecture is the process of designing access to information so that users can rely mainly on intuition to navigate quickly and productively and

“In essence, it enables access to a content by providing a systematic and primarily a visual approach to the organization of content and thus facilitates the quest for information. It contributes significantly to the information interface by the way that a user interacts with the content of a

## Rosenfeld and Morville

Rosenfeld and Morville (2002) give 4 definitions

- The combination of organization, labeling, navigation schemes within an information
- The structural design of an information space to facilitate task completion and intuitive access
- The art and science of structuring and classifying information on websites and intranets to help people find and use information.
- An emerging discipline and community of practice focused on bringing principles of design and user-centered design to the digital landscape

## Users, content, context

In designing an information system three aspects are considered (Rosenfeld and Morville, 2002):

- the context of the system: the business goals, organizational culture, technology, resources, and constraints;
- the contents the system provides: the documents, applications, services, metadata;
- the users of the system.

Though a good information architecture approach takes in account all those three aspects, we will focus on the users.

## Four areas of intervention

Rosenfeld and Morville (2002) identify 4 areas of intervention:

1. Organization: the way the information has been organized and structured.
2. Labeling: the lexicon to use to label the resources and the categories of the information system.
3. Navigation: the (hyper)tools provided by the system to the users to navigate into the system.
4. Search: the tools provided to the users to find the information.

# Organization schemes

## Exact schemes

- alphabetical
- chronological
- geographical

## Ambiguous schemes

- topic
- task
- audience
- metaphor
- hybrids

## Exact schemes

The file system of our computer gives us a good example of some exact schemes of organization. When we look at the directory of the file system we can sort it's contents by name, the date, the type of file (it's extension) or by file size dimension.

Another good example is the mail client. You can sort your in-box by date, by subject or by sender.

It is somehow easy to arrange an exact organization and you don't need to involve the users in sorting. Sometimes, however, it is difficult to find out what you are looking for. Ever experienced to looking for an email without knowing how to find it?

## Ambiguous organization

While ambiguous organization schemes require introduce a messy element of subjectivity, they more valuable to the user than exact schemes. simple reason why people find ambiguous organ schemes so useful: we don't always know what for." (Rosenfeld and Morville, 2002). Whenever Rosenfeld and Morville (2002) recommend using combination of the two schemes, considering w best for a particular audience.

## Organization structure

Rosenfeld and Morville (2002) identify three types of structures:

1. hierarchy; examples:
  - the directories like dmoz (the open directory project) or yahoo
  - the archive of a mailing list ordered by date (e.g., groups.google.com)
2. database (or list); examples:
  - the results of a query to a search engine like google or yahoo
  - the in-box of your mail client
  - the list of members of a yahoo group
3. hypertext (network); examples:
  - wikipedia

## Other structures

We can cite other two kinds of structures:

1. faceted classification: to take it short, is a multidimensional flat classification; examples:
  - [www.epicurious.com](http://www.epicurious.com): recipe collection
2. tagging systems and folksonomies: users or providers add some tags to the resources of an information system. Tags can be used to navigate resources; examples:
  - [del.icio.us](http://del.icio.us)
  - [flickr.com](http://flickr.com)

## Labeling

The title of a book is a label. The subject of a book is a label.  
The url of a website is a label (what do you expect from  
the site [www.sex.com](http://www.sex.com)?)

A label is a short piece of information that provides  
“information scent” about the contents of a resource.  
It can be a word or a short phrase (“contact us”) or an address  
(“envelope”).

## Labels

Labels are used everywhere in a web site. The name of the site are labels. The navigation system uses labels. The names of the groups or directories are labels. The contextual menu uses labels. "C" label, maybe the most infamous zero scent label. A good label should be short, informative, unambiguous. A labeling system should be coherent through the information system.

When possible, labels should follow conventions.

## The title of a resource

The most important label of a resource is the title. When you make a query on google, it gets you a list of documents. For each document the title, a short excerpt and the URL are shown. Users usually make their choice on which link to click on the titles of the documents.

The title should give an idea of the contents of the resource.

## The users

Information architecture has to deal with three context, contents and users (Rosenfeld and Mo...  
As we are proposing some methods for an user approach to the information architecture we will focus on the users.

What should we ask to the users? In my opinion user centered methods of IA should be able to answer the main questions:

- What does the users expect to find in a website?
- How do they expect the information should be structured and presented?
- What is their lexicon? What kind of labels and terms they use?

## Rapid IA Prototyping

Rapid Information Architecture Prototyping is a process for creating and testing IA based on user and business requirements (Sinha and Boutelle, 2004).

Steps:

1. **stakeholder analysis**: to understand business and organizational context
2. **free listing**: to explore the domain and to capture users' lexicon.
3. **q-sort** if needed: to discover the important items for the users.
4. **open card sorting**: to discover the classification of the users.
5. **closed card sorting**: to categorize the items into a taxonomy.

## Free listing

In IA the free listing can be used for two purposes:

1. find out the contents, the scope and the boundaries of a semantic domain;
2. elicit the implicit lexicon of the users;

How to conduct a free listing?

In a free listing questionnaire, the participants are asked to name all the items they can remember of a domain or category.

## Examples

- Please name up to 10 animals in the category mammals [fish/reptiles/birds].
- What do you expect to find on the web portal of Trento University? Which information? Please list the pages you wish to find on the web portal of the university.
- Where would you like to go for holiday?

## Analysing results

If you test an adequate number of participants, gives you a list of the items of a semantic domain should represent the tacit expectations of the users. The words used. Those results can therefore be the labeling system of the web site.

If you use an online tool (as we do) you will collect garbage, some typing errors and many synonyms. You need to clean up the results to get a coherent and usable

## Ranking

The items obtained can be ranked by frequency and individual rank. One simple way is to give 5 points to the item that appears at the first place of an individual, 4 points to the second and so on, and 1 point to the 5th to 10th. Frequency and rank order usually correlate well (Bousfield and Barclay, 1950). Free listing frequency and prototypicality evaluation correlate as well (Bussey, 2005).

The resulting rank can give us a hint on the importance attributed to the items by the participants.

## Integrating and selecting t

The list obtained from the free listing has to be combined with (and integrated with) the contents of the existing site and with a similar list provided by the stakeholders. The final list will be the basis for the card sorting. If the number of items in the list is over 60-80 items it is preferable to reduce the list to the 60 most important items for the card sorting. Probably some of the items obtained by the free listing were not planned to be included into the informative architecture. The content manager has to decide if to integrate them. This decision should be based on the importance the stakeholders attribute to each item and the costs of the integration. Anyway those items should guide the stakeholders in the planning of the future versions of the site – horizontal expansion (Sinha and Boutelle, 2004).

## Similarity matrix

There are two ways to indirectly compute a similarity matrix from a free listing (Sinha, 2003):

1. Co-occurrence: calculates how many times pairs of items co-occur.
2. Average rank distance: the closer two items are in free-lists, the more similar they are supposed to be perceived.

## Other possible analysis

There are many other ways to use free-listing data (see Galletta et al., 2003):

- estimate the domain familiarity of each participant;
- identify the “outliers”: respondents who, for some reason, might perceive the domain in a completely different fashion;
- ask participants to generate multiple free-lists; the first free-list is used as starting point for the second;
- compare the content and boundaries of two different domains.

## What is the cardsort

The Card Sorting is a user-centered design method to improve the findability of a web site.

The sorting task requires that the users group a set of labelled cards in coherent sets.

Participants could be asked to suggest a label for the sets they formed.

The cardsorting is an effective tool to elicit and analyze implicit mental models of the users. This helps information architecture experts to better organize the site's information that will be easier to find and to use, thus improving the quality of the site.

## When to use card sorts

Sorts are appropriate when the emphasis is on the categories which people use (Rugg and McGeorge). Using card sorting we can discover:

- the criteria participants use to categorize and to search it;
- the expected structure of the information;
- the differences among different groups of users;
- the labels of the categories, in the terms of the user's syntax.

## How to prepare it

First of all we need to prepare the cards with the labels for each item. We need to evaluate the appropriate labels, that have to be understandable; it could be necessary to perform a preliminary labelling analysis. It could be useful to have some boxes where the cards can group the cards; the number of the boxes should be the number of maximum categories we wish to create (if we already have decided the labels for the groups of cardsorting) we should label the boxes.

## Running the test

- At the beginning we have to inform the participant about the aim of the experiment;
- we give him the instruction on how to sort the cards. It is important to avoid to give him information that could influence his choices;
- we shuffle the cards and we present them to the participant;
- in few minutes the participant have to put the cards into the boxes.

## Statistical analysis

The data collected can be analyzed using the hierarchical cluster analysis.

The first step is to calculate a  $n \times n$  proximity matrix where  $n$  is the number of items.

The proximity matrix can be produced by calculating for each slot  $i,j$ , the number of times where the items have been grouped together.

The cluster analysis can produce a tree graph of hierarchical classification.

## What is netsorting

Netsorting is a web based cardsorting tool.

It is based on java servlets and postgresql database on the server side.

It uses plain html and javascript on the client side.

Netsorting allows to analyse the data implemented

the statistical methods most used for the card sorting.

## Why using netsorting

There are many benefits in using a web based tool:

- it is easier to recruit participants with an cardsorting;
- if you are using netsorting to restructure a site you can recruit the **real** users in your
- the recruitment can be conducted via e-m.
- no data-entry is needed;
- the analysis of the data can be performed

## Features of Netsort

The following features are implemented in netsort

- Free listing
- Likert scales; used for
  - importance evaluation: participants have to rate the importance of each item;
  - prototypicality evaluation: participants have to evaluate how each item fits into a category;
- closed card sorting: participants have to sort items into some specified (labelled) groups;
- open card sorting: participants have to sort items into some unlabelled groups; after that they have to label the groups used and to indicate the order of importance in classifying the items;

## hierarchical cluster analysis

- computation of the proximity matrix of the classification;
- visualisation (as an html page) and export (as a file) of the proximity matrix;
- computation of the hierarchical cluster analysis from the proximity matrix;
- online visualisation (through a web page) of the dendrogram of the hierarchical analysis.
- an interface for the manipulation of the dendrogram (status: alpha)

## Analysis of the participants

- an algorithm to group the participants in based on the different conditions (for example versus offline, open versus closed card sorting, random assignation; this function have been used in experiments we are conducting using Netscape);
- computation of the proximity matrix for each group created;
- computation of the correlation index among the proximity matrixes of the groups, based on the method used by Tullis and Wood (2004);
- computation of the correlation index among the proximity matrixes for each item of the cardsorting;

# Thank you

For further infos:  
stefano@bussolon.it  
<http://www.cardsorting.info>

## References

- Bousfield, W. and Barclay, W. (1950). The relationship between order and frequency of occurrence of restricted associative responses. *Journal of Experimental Psychology*.
- Bussolon, S., Ferron, M., and Del Missier, F. (2005). On-line categorization and card sorting. In *Cognitive processes in ergonomics and human computer interaction*. 7<sup>th</sup> Alps-Adria conference in psychology, Zadar, (Croatia).
- Rosenfeld, L. and Morville, P. (2002). *Information architecture for the World Wide Web – 2nd ed.* O'Reilly and Associates, Inc., Sebastopol, CA, USA.
- Rugg, G. and McGeorge, P. (1997). The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems*, 14(2):80 – 93.
- Sinha, R. (2003). Beyond cardsorting: Free-listing methods to explore user categorizations. Technical report, Boxes and Arrows.

Sinha, R. and Boutelle, J. (2004). Rapid information architecture prototyping. In *DIS '04: Proceedings of the 2004 conference on Designing interactive systems*, pages 349–352, New York, NY, USA. ACM Press.

Tullis, T. and Wood, L. (2004). How many users are enough for a card-sorting study? In *Proceedings UPA '2004*, Minneapolis, MN.