Information Architecture

Tobias Zimmermann

Abstract: This paper introduces to the field of information architecture. It starts with an problem oriented view on cognitive overload followed by a short introduction and definition of information architecture. I will describe different types of user information seeking behavior to understand how an information architecture must support information retrieval. A short introduction into a user centered development process of an effective information architecture follows. The main part of this paper discusses the components of information architecture, how organization, labeling and navigation systems are set up. Further components such as search system and controlled vocabularies for an semantic search are described. This paper finishes with an short outlook on the implications of a Next-Generation-User-Centered-Information-Management to information architecture.

Cognitive Overload

In today's information decade 'information overload' is an buzzword to describe the problems resulting from the incredible amount of information. But information overload is only a facet of the problem. But are these problems only cause by the simple presence of information quantity. The research on cognitive overload differentiate between several facets of this problem. Kirsh [Ki00, 23] identified four drivers of cognitive overload:

- too much information supply,
- too much information demand,
- the need to deal with multi-tasking and interruptions,
- inadequate workplace infrastructure.

Kirsh argues that the oversupply of pushed and pulled information is caused by user information retrieval behavior as well as the heavily increased amount of available and accessible information. He implies a exponential growth in overall information quantity but assumes an only linear growth in (high) quality information quantity (see figure 1).

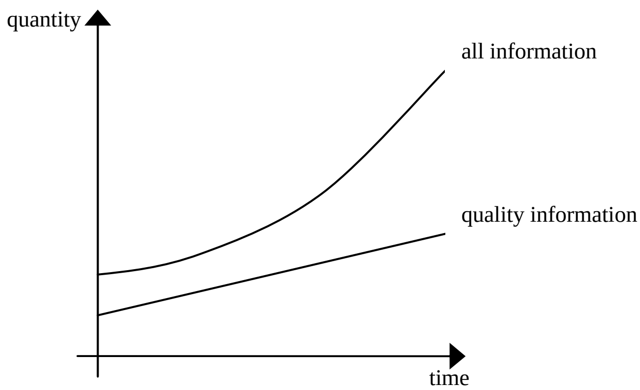

figure 1: quantity of all and quality information, adopted from [Ki00, 26]

Kirsh shows us that the problem of information overload is not simply caused by the increase of information amount but by other factors like the increased number of decisions of knowledge workers, the need of multi-tasking, the increased frequency of interruptions, the need of time efficiency and the lack of adequate work environments. People suffer from information anxiety with negative influence on their health [Ki00, 22].

Defining 'Information Architecture'

An information architect tries to address the problem of cognitive overload by adding adequate structures, labels and browsing aids to websites and software applications in order to improve usability. In fact an adequate information architecture may optimize the digital work environment which is the information space. Users should easily find the information they requested, without being overloaded with too much information. They should work in the information space in an intuitive way and should be able to recover their work after an interruption.

There is no clear and generally accepted scientific definition for *information architecture*. One approach of the Asilomar Institute for Information Architecture describes information architecture as follows:

1. *The structural design of shared information environments.*
2. *The art and science of organizing and labeling web sites, intranets, online communities and software to support usability and findability.*
3. *An emerging community of practice focused on bringing principles of design and architecture to the digital landscape.*
4. *A structure based on the patterns inherent in data that allows users to accomplish their goals.*

Often mentioned in this context are practices and task which are not information architecture. These boundaries should help to understand the work field information architecture [RM02]. It is **not**

- graphic design
- software development
- usability engineering

But the development of an information space goes hand in hand with these disciplines and there are several intersections.

User information needs and information seeking behavior

To design an adequate information space it is necessary to understand the users' requirements and their information retrieval behavior. We can identify three different *information needs* which cause three different information retrieval behavior. A good information system/space must support all of them to satisfy users' demands.

The most structured proceeding in retrieving information is to know how the information is called and described and where to look after it. We call this sort of information retrieval "*known-item seeking*". A good example for this action is to look up a person's phone number and we know the name of the person and where his phone number is listed. One single search query should satisfy the user's information need.

A more unstructured information retrieval is performed if the user does not exactly know what he is looking for. In fact he is aware of the context and where he may find the information but not exactly. He is querying and browsing through the information space iteratively. The process of searching is at the same time a process of learning about the information he is looking for. This results in a larger amount of information about the topic. This information retrieval is called *exploratory seeking*.

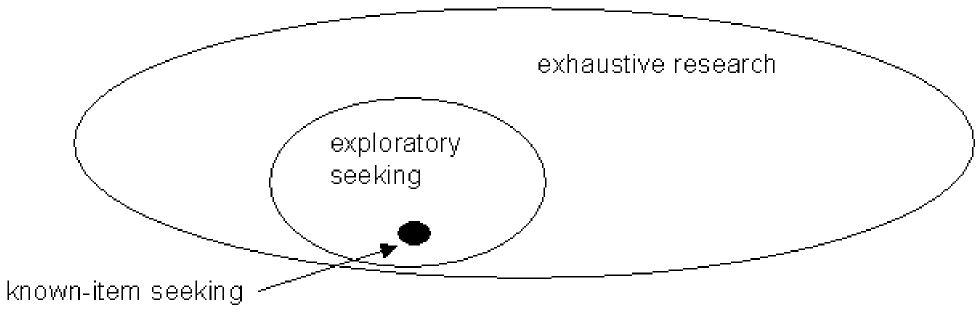

figure 2: Three types of information needs

The last type of information retrieval is called *exhaustive research*. The user wants to all information he could get on a certain topic. He will perform several search queries with different search terms like synonyms and broader and narrower search terms. This may result in the largest amount of information and there is no single right answer on the information demand. An example would be an scientist who is investigating on a topic and desires all information publicized on his topic.

Effective Information Architecture

Effective information architectures can be developed in a time consuming and complex project. In fact it is a multifaceted discipline. But on the other hand the consequences of an ineffective information architecture may result in cost which are hardly to quantify but agreeable to high. These efforts may be

Costs of …

- ... finding information, if every employee may spends 10 minutes daily less in seeking his required information on the corporate intranet/information system,
- ... not finding information, if available information are obtained several times from an external information provider,
- ... redesign and improvements, to increase the quality of an information space and fulfill user requirements,
- ... maintenance,
- ... training, to help frustrated employees in using the corporate intranet,
- ... lost customers, who got lost in the company's online shop and purchases the product at an competitor's shop,
- ... lost reputation and brand value,
- …

An effective information architecture relies on the interdependencies of *context*, *content*, and *users* (see figure 3). These interdependencies gives an information architecture a unique identity. The three circles of information architecture define the environment or "information ecology" in which the information space "lives" and grows over time.

Information spaces exist in a particular context, which is defined by the business or organizational *context* [RM02, 24]. The information architecture reflects it's business goals and strategies, as well as internal politics, funding and ownership. These cultural aspects or business context has to be part of the information architecture to be accepted and adopted by the organization. Technology and processes are as well part of this context. The information architecture must fit and support internal technologies and processes to be an effective part of the organization.

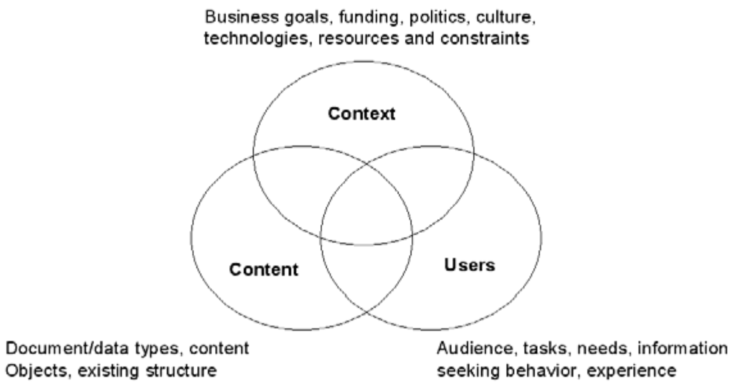

figure 3: Three circles of information architecture

Existing and future *content* of the information space is the next facet influencing the information architecture. Content summarizes all the documents, metadata, applications and services that users need to find in the information space [RM02, 24]. While the information architecture should enable adequate access to existing and future content it is defined by the content types and data objects it should represent. It is a matter how content is stored and organized.

The last facet *users* is the most crucial one. It is obvious that an information space has to address its' users and customers. An information space is a tool to satisfy their needs. If it does not so users will choose an other tool or simply get frustrated if there is no alternative. An effective information architecture has to deal with users' needs but there is no uniform 'user'. User differ in experience with information spaces, in task they have to perform and in their information seeking behavior. To implement an effective information architecture we have to investigate exactly who are those users and their characteristics.

Developing an effective information architecture

Within this information ecology we should follow a formal process to design and develop the information architecture. This process is split in two tracks as shown in figure 4. It starts by 'Defining the audience'. In this step we should clearly define for whom we are designing the information architecture. Who are the users and customers? To understand clearly the target audience we have to investigate through user research the

- needs,
- desires,
- abilities and
- methods

of the target audience [Ad01]. A useful way to summarize those findings are *personas* and *scenarios*. Persona is a fictitious person with concrete characteristics and archetypal qualities of the audience. They are stereotypes to illustrate the different users. Several personas and scenarios of using the information space may give a good impression of the target audience.

Once we identified and understood the target audience the process splits up in two separate tracks, the *mental model* and the *conceptual model* (see figure 4). The mental model represents the content and structure of the information space "… how the users think about and approaches their tasks and goals." [Ad01]. This model helps us to understand the tasks and needs users have. The other track results in the conceptual model. It represents the information space and its' content from the information architect's point of view. It gives us a structure and patterns of the content available on the information space. In this track the development of an organizational, navigation and labeling system is started.

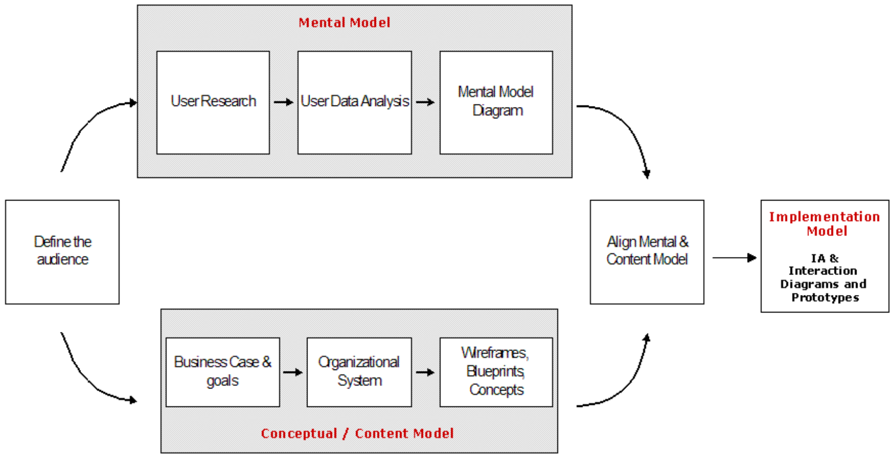

figure 4: Process of developing an effective information architecture, adopted form [Ad01] and [Ch01]

The most difficult task is to align these two tracks together in one single and comprehensive model. The mental model should sharpen the view on a user-centered information architecture besides the conceptual model which stresses the organizational issues and the business goals. Once we aligned these two models we are on the way to achieve the ultimate information architecture design goal to develop an “information architecture that corresponds to our users’ mental models and also meets the business goals” [Ad01]. A gap analysis is a useful tool to check whether the alignment fits the requirements. It shows us if users expect the content we are providing on the information space and if the organizational system orders the content in comprehensive clusters. Content gaps are hints we have to add additional content or to optimize the navigation and organization system. If both models do not fit the alignment we have to step back and set up a better conceptual model.

Information Architecture components

To understand the variety of concepts and task in developing an information architecture it is useful to elaborate and describe the components of an information architecture. There are several classifications schemes of them in literature and [RM02] gives a non exhaustive overview. I’ve decided to follow the scheme of [RM02] because it shows the reader an easy and comprehensive insight to the basic principles of information architecture.

Organization Systems

Content amount and diversity of today’s information decade makes it necessary to organize information so that users are able to retrieve them later in an efficient way. An information architecture enables this by classifying and cataloging the content and labeling these structures (see next chapter for labeling systems). Organization Systems are composed of *organization schemes* and *organization structures* [RM02].

Organization schemes define the shared characteristics of content items, similar content items are arranged in logical groups or categories. The scheme may also define a certain sorting of content items such as alphabetical, chronological or geographical order. These schemes are *exact organization schemes*, a content item can only be member of a unique logical group (e.g. a city relates to exactly one county). Other organization schemes, called *ambiguous organization schemes*, do not have this restriction. Organization schemes of this type may be by topic, by task, by audience or by metaphor (e.g. a news article about the latest economic decisions by the political leader may be member of the topic category “politics” and as well of “economics”, see also figure 5).

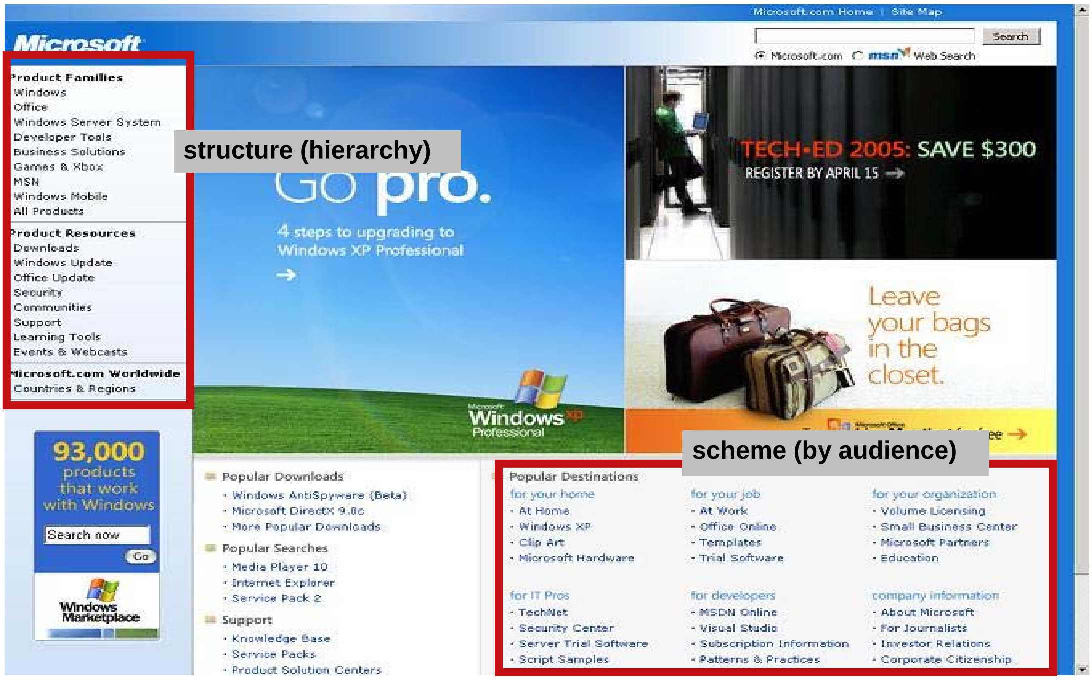

figure 5: microsoft.com website with organization scheme and structure

Organization structures define the relationships between those groups or categories. Most commonly such relationships are hierarchical (parent-child-relation, see also figure 5). But also database-models, similar to entity-relationship-diagram, used by content management systems, are organizational structures. By bringing both approaches together we will get a network of interwoven hyper textual connections between our content groups. These schemes may represent personal associations or simply different perspectives on the content for different target audience. But the risk of getting lost due to complexity arises.

In case of hierarchical structures this organization system reflects a taxonomy and allows us to categorize content. The parent content group contains broader content items, the child category a narrower scope of content.

To categorize existing amounts of information the card sort practice achieves good results. All content titles are written on small cards and several users and information architects have to arrange them in an expedient manner. The resulting structure may be a good user-centered organizational structure.

Organizational systems together with labeling systems are the basis or anatomy of an information space. Its' structures can be found again in the content structure and navigation structure. In practice some well defined *information architecture models* or *information architecture patterns* have been developed.

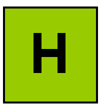

figure 6: **All-in-one model**, adopted from [Sc04a]

The **All-in-one model** is the most simple structure model. All content is allocated on one single page. It fits best for small websites with a small amount of content. The user can overview all content on a single sight.

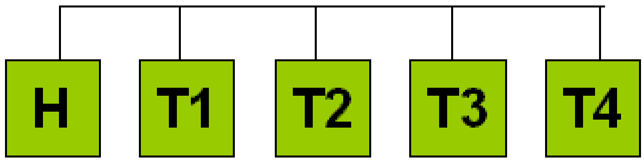

figure 7: **flat pattern / monocline grouping**, adopted from [Sc04a]

The **flat pattern** or **monocline grouping** fits best for simple websites which have only a few standard topics (e.g. Home, Contact, Products, About us, …). Every page is accessible from all others. They are arranged as peers.

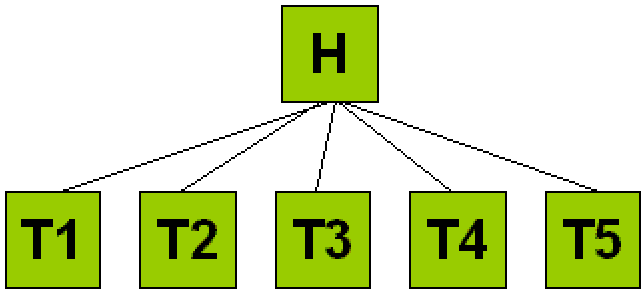

figure 8: **Index pattern**, adopted from [Sc04a]

The **index pattern** is a similar model as the flat pattern but there is at least one index site or the homepage itself is organized as an index page. This pattern is soliciting if the content is similar to a directory with a certain sorting order (e.g. names directory). It may provide a quick access to a huge amount of information, if the user is performing a known-item-search.

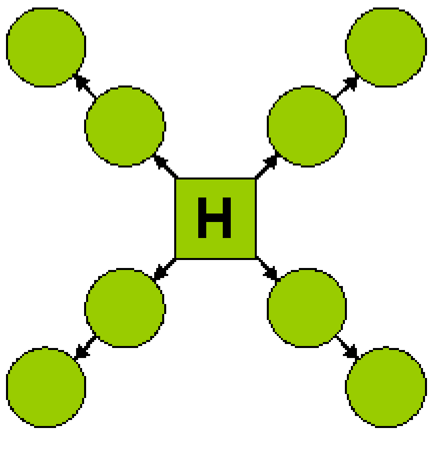

figure 9: **Hub-and-Spoke model**, adopted from [Sc04a]

The **Hub-and-Spoke model** provides useful access to several distinct and linear workflows starting all at a single point. Think of a web-interface of an email service provider. The starting point is the users inbox from where he can start several functions as read email, write an reply, create new mail, … Those workflows contain several steps and will always return the user to his inbox as a starting point.

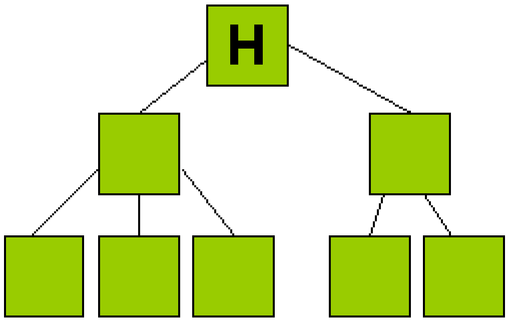

figure 10: **Strict hierarchy**, adopted from [Sc04a]

Within a **Strict hierarchy** model structure all sub-pages (child nodes) are only accessible via its parent pages. This pattern represents best a given taxonomy or a corporate organization scheme. The overall constraint is the strict 1:n relationship between parent and child content items (web pages). Every parent item may have several sub items and every child item has one distinctive parent item.

But beware of this restriction. It is possible that this not common sense for all users. A corporate product portfolio is often structured in a strict hierarchy but this scheme may not be as comprehensible for your customers as for your marketing staff. There for it may be a better way to represent a strict hierarchy within a different information architecture model.

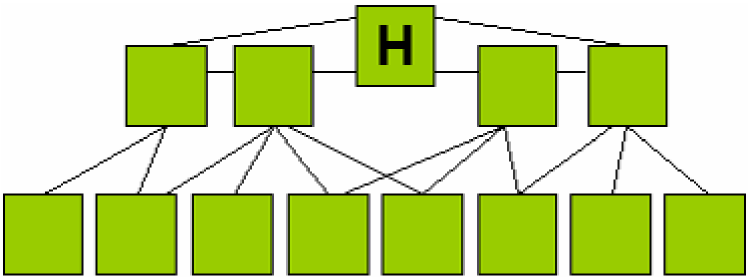

figure 11: **Polyhierarchy** or **multi-dimensional hierarchy**

A **polyhierarchy/multi-dimensional hierarchy** is a more flexible variation of an strict hierarchy. The 1:n constraint in parent-child relationships is replaced by an n:m relation: A parent node may have several child node and a child node may have also several parent nodes. This causes the effect that a content item may belong to different logical groups. Customers may have different access paths to content items, they can find a product in a online-store in different categories with less risk that he looses sight of the desired product. On the other hand this organization system needs the highest effort in maintenance and care.

Labeling Systems

Labels are given names to represent larger chunks of information. It is somehow the visual representation of the organization scheme and structure. By labeling systems we understand the complete set of labels in an information space together.

Labels can be either textual labels or iconographic labels. Some examples for textual labels are hyperlinks, headings, navigation options, menu titles and index terms. Iconographic labels use the visual aspect of an icon or graphic to represent content. They are often used as layout and design elements. Labels use the audience expectations and understandings of their meaning to summarize and visualize the content behind it. But anticipating these expectations are the major pitfalls in designing such labeling systems. Icons can represent more complex functions and express more content shorter than text but also can easily be misunderstood. “A picture is worth than thousand words, but which thousand?” [RM02, 91]. A similar problem occurs if vocabularies used for textual labels are organizational jargon, specialized terms used by the employees but not by the users. [RM02] give us four major rules in designing labeling systems:

- develop consistent labeling system, not just labels
- use narrow scope whenever possible
- be aware of different “languages”/perceptions
- avoid noticeable gaps in the labeling system

Consistency tables can be used to avoid inconsistent labeling systems especially at the navigation system. These tables list all navigation links with labels and target site names. This gives us a good overview if we used the same and comprehensive labels on the entire information space. Sources for labeling systems may be comparable or competitive information spaces. May be there is a industry quasi standard users already get used to. Also controlled vocabularies or thesauri can be used to obtain a comprehensive and consistent labeling system.

Navigation Systems

The most crucial [RM02, 107] information architecture component for users is the navigation system of the information space. It makes the content accessible to the target audience and combines the visual and usable functionality of the organization and labeling system. It is regarded as the most crucial component to the user because it will get a source of frustration if the user gets lost in the information space. And customers who got lost are in most of the cases lost customers. They simply click away to a competitor website where the navigation system is more usable.

We differentiate 3 types of navigation systems:

- embedded navigation systems,
- supplemental navigation systems and
- advanced navigation approaches.

Embedded navigation systems consists of *global navigation*, *local navigation* and *contextual navigation*. Embedded navigation systems are shown within the information space together with content. Figure 12 shows us the most popular navigation scheme for embedded navigation systems.

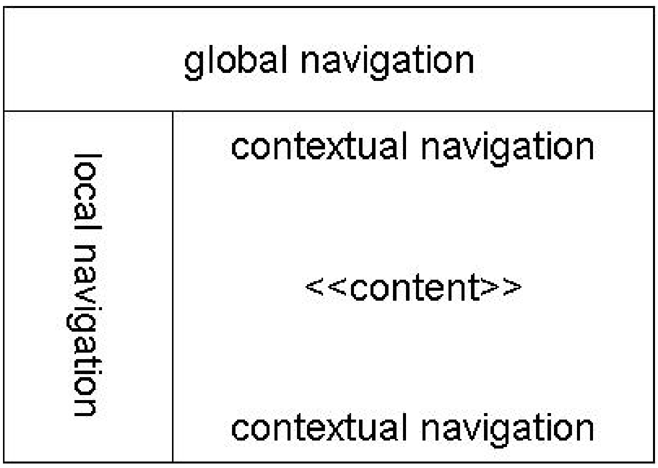

figure 12: Popular navigation scheme layout. Adopted from [RM02, 119]

The *global navigation* enables the user to browse through the main areas of the information space, it is unique and identical on most of the sites. The *local navigation* enables the user to navigate in the actual area of the information space, it may change in different areas. The contextual navigation is embedded into the content, it consist of hyperlinks or breadcrumb navigation. It has the most distinctive granularity.

The *Supplemental navigation* is mostly situated beside the content and may consist of sitemaps, indexes and guides or wizards. They provide different ways of accessing content directly in the information space. They are organized different to the overall organization system as an alphabetical order or birds eye view to the content.

These two types of navigation systems can be found on most information spaces and especially at websites. [RM02] gives us six principles of good navigation systems to consider:

- Let me know where I am all the time!
- Clearly differentiate hyperlinks from content!
- Let me know clearly where I can go from here!
- Let me see where I've already been!
- Make it obvious what to do to get somewhere!
- Indicate what clicking a link will do!

In fact these principles are common sense but often not satisfied in information spaces especially at huge websites.

While the first two types of navigation systems are necessary and widely accepted they may be not sufficient in some cases. *Advanced navigation approaches* aroused in some niche markets. But because of their narrow scope in usability, lack of user experience and their difficulties to implement in a efficient manner they did not become generally accepted navigation systems. Examples for such approaches are personalization, customization, visualization and social navigation. Personalization tries to present navigation options to the user based on his past behavior and information access. It works without direct user interaction. A good example is the amazon.com recommendation system presenting similar books to the user as he already bought. In contrast customization gives the user the full control over the navigation design. A reader of a news portal is able to customize the content areas presented to him on the entering page. Further approaches such as visualization try to represent the information space with metaphors like information rooms being capable of bearing associations and relations of the user to the content. Cushion view diagrams and self-organizing maps give the user an impression where to find content and presenting an additional dimension (such as activity, amount, size, …) to him. At least there is social navigation to mention as a alternative approach for navigation systems. These navigation approaches try to retrieve the value of a certain content item for one user by observing the behavior of other users. Such approaches could be manifested as "most common downloads" or bestseller-lists in an online-shop. Other users can actively recommend content item by ratings to show which content is valuable. An other approaches are "folksonomies". Users give personal tags to their content to organize their personal items. By browsing other users' personal tags the user may find related content. A good example for this approach is the photo library flickr.com.

These approaches actually play in present a niche role. They are not widespread accepted and do not fit on every information space. Such navigation approaches enable relations between content items which are not planned or predictable. They build up a net of interwoven content items and provide a way through the information space besides a strict and planned organization system.

Search systems

The functionality of searching a information space to find certain content is an additional component of information architecture. It is additional in the way that at first the other information architecture components should be well designed and finally a search system could be implemented. A search engine may pay of for an information space in the following cases:

- at huge amounts of information,

- at fragmented sites,
- with highly dynamic content,
- when users expect it.

But a search system is not the silver bullet to improve the navigation in an information space, the search functionality can not help users when the usually get lost in the information space. This is clearly an lack of the navigation system which can not be balanced by a search system.

But a search functionality may help users to retrieve information and content easier and quicker from the information space. Search systems usually use indexing algorithms to retrieve content quickly which matches the users' query. The essential benefit for the user is the search result and its' ranking. Search results can be presented in a ranking by relevance or popularity, in a sorting such as chronological, alphabetical or geographical or grouped and aligned by clusters of similarity. The search result presentation depends on the nature of the information space.

Metadata and Controlled vocabularies

At least there is to mention an invisible part of the information architecture. Metadata and controlled vocabularies are used to empower the ability of information retrieval at the information space. Metadata are invisible attributes or descriptions for content. Examples for metadata are attributes like author, keywords, language, date of publication, …. Typically search engines evaluates those metadata to provide further information in the search results or use it in a advanced search (e.g. find all books written by author xy).

In a next step controlled vocabularies are used to perform a semantic search. A semantic search tries to retrieve all the information the user meant by his query, not only the exact syntactical match of his search string. The search system uses the semantical relations between the search string and controlled vocabularies. Controlled vocabularies are manually cared term lists and relations between those terms. Synonym rings lists all terms with and equivalence relation (e.g. pda = handheld pc) so that a search algorithm is able to retrieve content with equivalent information. Authority files are lists of preferred terms, sometimes in combination with synonym rings. Classification schemes bear hierarchical relations between terms in form of broader an narrower terms ('beagle' is a 'dog'). The last step is the thesaurus. It includes equivalence, hierarchical and associative relations together and is capable to find related terms. See figure 13 for an overview of these vocabularies and their semantical relationships.

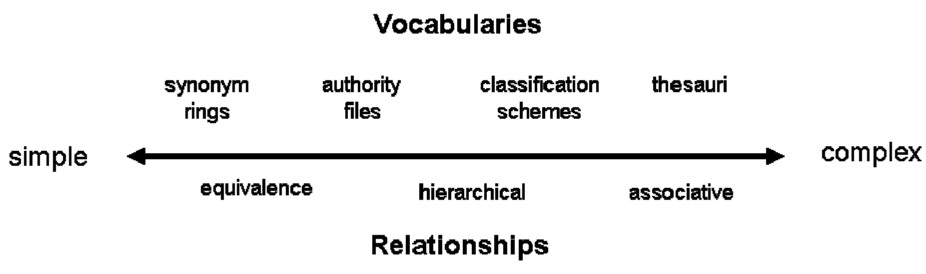

figure 13: Controlled vocabularies and their relations

Information Architecture and Next-Generation-User-Centered-Information-Management

What implication for information architecture come along with Next-Generation-User-Centered-Information-Management (NGUCIM)? We think in future the role of huge static information spaces will diminish and Personal Information Management will arise. User will organize their information in their own manner, they want to associate information with content in a more natural way. This means that we need more flexible organization systems than simple hierarchical patterns and information representation with files and folders. The result of huge enterprise knowledge management / content management systems were employees being confused by their own companies labeling systems making knowledge sharing a hard work.

NGUCIM and personal information management implies that information architectures are set up by the users their selves especially in the area of organizational and labeling systems. By sharing these systems with others there is a possibility of getting a conformity of widespread accepted systems instead of centrally build and misunderstood systems. We think that the principles of 'folksonomies' can contribute a lot in this field.

But this implies a change in the actual concepts of navigation and visualization systems. We need highly capable navigation systems to express those folksonomies and to see the 'world' around the personal information space. Users must be capable to browse in a interwoven network of associations between internal and external information without loosing the oversight of their and others' information.

References

[Ad01] Adaptive Path : Designing the Complete User Experience,
url: http://www.adaptivepath.com/presentations/complete, accessed: 21.03.05.

[Ch01] Chak, A. : Effective Info Architecture.
url: http://www.webtechniques.com/archives/2001/10/chak, accessed: 26.03.05.

[Ki00] Kirsh, D. : A Few Thoughts on Cognitive Overload. Intellectica, 2000/1, 30, pp. 19-51,
url: http://interruptions.net/literature/Kirsh-Intellectica00-30.pdf, accessed: 28.03.05.

[RM02] Rosenfield, L. ; Morville, P : Information Architecture for the World Wide Web.
O'Reilly, 2002.

[Sc04a] Scratch Media: Navigation, 2004,
url: http://www.webdesignfromscratch.com/navigation.cfm, accessed: 28.03.05.

[Sc04b] Scratch Media: IA models, 2004,
url: http://www.webdesignfromscratch.com/ia_models.cfm, accessed: 28.03.05.