Chapter 1

User Centered Information Retrieval,

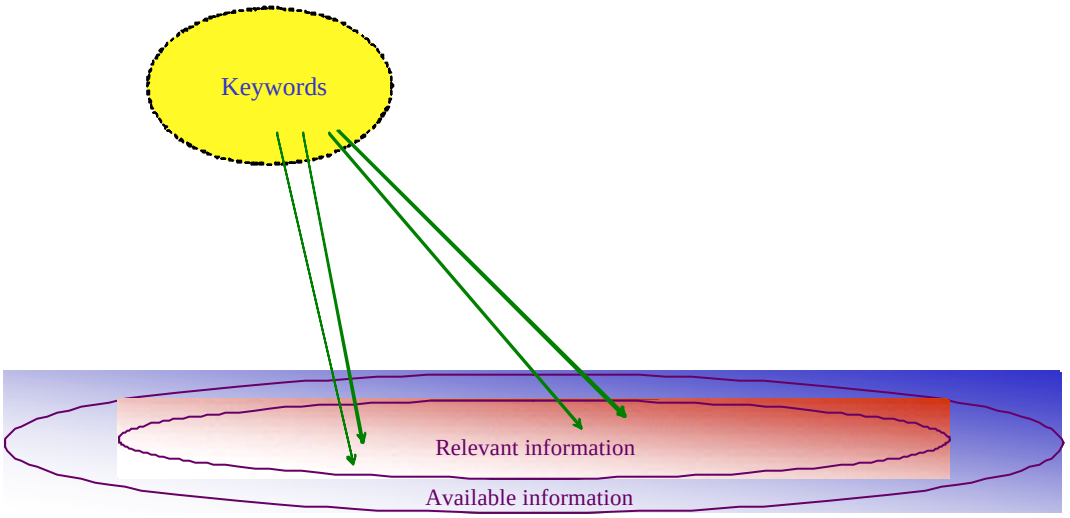
development method

Orientation

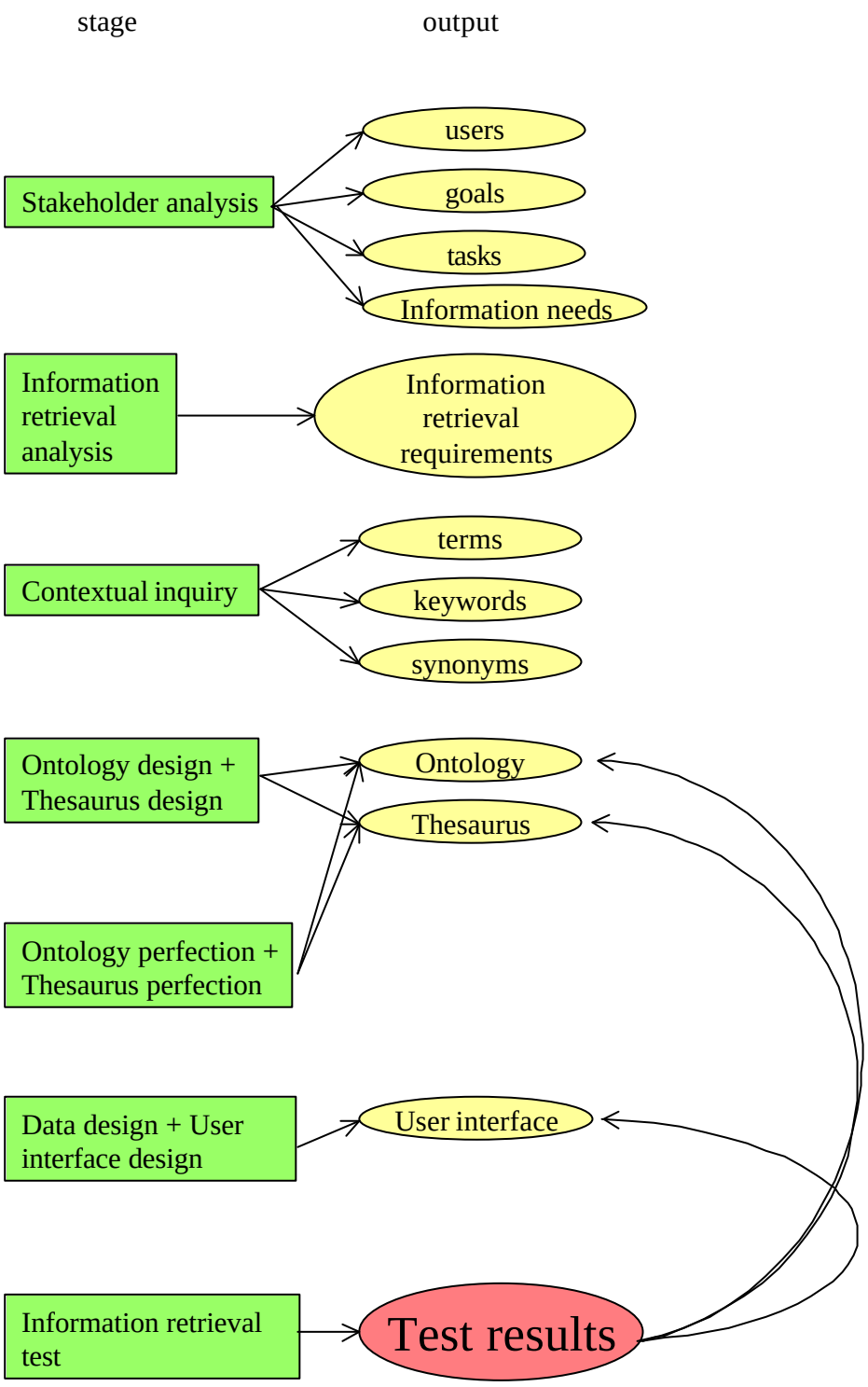
This chapter assumes the following:

- You want to make information findable for your readers
- You want to use metadata to make your information more findable
- You understand that keywords can be subject-oriented or reader-oriented
- You know how to get and write down the function of information

In this chapter we will describe the method that enables you to make information findable, using metadata.



User Centered Information Retrieval, Development method



User Centered Information Retrieval, Development Method

Stage	Purpose	Output	Explaining
Stakeholder analysis	Find and describe the people that need to find information	Users	Background, experience and function
		Goals	Why finding the information is important to them
		Tasks	The tasks that need to be supported by information
		Information needs	The information that is necessary to perform the tasks
Information retrieval analysis	Determine the minimal and the desired requirements for finding information In the Information Retrieval Test these requirements need to be met	Information retrieval requirements	• Ratio: $\frac{\text{successful searches}}{\text{search attempts}}$ • Number of search refinements before finding • Time before finding • Precision: how much non-applicable information is allowed to be found • Recall: how often is information not found although it is present • How applicable the information must be
Contextual inquiry	For each reader, find the 'language' of the reader: "How are things called in the context of the reader?"	Terms	Words and what they mean in the context of the reader
		Keywords	The terms that the reader uses to search
		Synonyms	Words that mean the same to the reader or words that are called differently by another reader
Ontology design plus Thesaurus design	A first draft of the metadata that will be used to find information	Ontology	Functional and subject-oriented terms that enable the reader to find information. For each term write down what it means to the reader
		Thesaurus	Containing the synonyms for the terms in the ontology
Ontology and Thesaurus perfection	Bottom up approach: check whether all information can be assigned to one or more terms in the ontology	Improved ontology	Terms that need to be added to the ontology because they were 'forgotten' in the initial ontology
		Improved thesaurus	Synonyms that were 'forgotten' in the initial thesaurus

User Centered Information Retrieval, Development Method Continued

Stage	Purpose	Output	Explaining
Data design and user interface design	Take a decision on the structure that the metadata will follow The choice of the structure of the metadata also determines which User Interfaces can be applied	Data design	Choice between: • RDF • Topic Maps • OWL • XFML
		User interface design	The user interface that supports the reader to search for information, refine his search, browse through the search result and retrieve the information
Information retrieval test	Test the requirements from the "Information Retrieval Analysis"	Test results	• Number of times information was not found • Recall: Number of times information was not found although information was present • Precision: how much non-applicable information was found • Number of search refinements necessary before finding the information • The time to find information • How applicable was the information found? • Keywords used, a further perfection of the ontology • Synonyms used, a further perfection of the thesaurus

Stakeholder analysis

1. Target audience: who are you helping by improving the findability of information? Be creative here. Most of the time more people benefit than you may think at first.
2. Goals: what exactly is the benefit? Why do these people want to have (better) findable information? What would go wrong if information was not (that well) findable?
3. Tasks: as specific as possible, describe the user scenario's that describe the occasions when the reader will benefit from the information
4. Information needs: a description of the information that is used during the tasks

Information Retrieval Analysis

Requirement	Explanation
Ratio: successful searches / search attempts	What is the minimum and desired ratio when you divide 1. How often the reader indicates he has found the information by 2. How often the reader searches '0' means information is never found '1' means information is always found
Number of search refinements before finding	What is the minimum and the desired number of times the reader • Makes a choice or • Types in a search
Time before finding	The minimum and the desired time between 'the reader understands the task / question' and 'the reader indicates he has found the answer'.
Precision: how much non-applicable information does the reader find?	The minimum and desired number of times the reader retrieves information but decides that the information does not contain the information he needs.
Recall: how often is information not found although it is present?	The minimum and desired number of times the information is not found although the information is present
How applicable must the information be?	The minimum and desired rating you want the reader to give when the reader indicates how applicable the information is?

The customer decides for every task, for every requirement, a minimum value and a desired value. This results into an *Information Retrieval Requirements Table*. Here you see an example:

Task	Hit ratio		Refinements		Time		Precision		Recall		Applic.	
	Min.	Des.	Min.	Des.	Min.	Des.	Min.	Des.	Min.	Des.	Min.	Des.
1.												
1.1.												
1.2.												
1.2.1.												
2.												
3.												
3.1.												
3.2.												
3.3.												

Min. = minimum value Des. = desired value

Contextual inquiry

In the contextual inquiry you try to find out the terms the reader uses. Some of these terms will be reader-oriented, some of these terms will be subject-oriented. Find all. Make note of

- The terms used when the reader looks for information (e.g. terms used in a question). These are the keywords.
- Different terms that mean the same thing to a reader or the same thing is called differently by another reader. These are the synonyms.

Ontology design

An ontology is the collection of the terms plus for each term the definition. The ontology contains the terms the reader will see when browsing. Using the task information found in the Stakeholder Analysis and all the information from the Contextual inquiry, write down the list of terms and their definitions. Do not put too many terms in your ontology, less is better (producing a more effective set of metadata).

Thesaurus design

A thesaurus is a list of synonyms for your ontology. Each term in the ontology can have one or more different terms in the Thesaurus. The synonym itself can also be in the ontology. For example when another reader uses a different term when he means the same. Two terms mean the same when they refer to the same definition.

Ontology and Thesaurus Perfection

Take the information you want to make accessible. Use this information for a bottom up approach:

- make sure you can link every piece of information to at least one term in your ontology. If an item cannot be linked, it may be your ontology is not complete.
- retrieve the reader-oriented terms (when to use this information?) and the subject-oriented terms (what is the information about?) from this information and if necessary, add them to your ontology
- look for synonyms in the existing information

Data design

Up till now you have been constructing an organization of the information that supports your target audiences. But also your metadata (ontology plus thesaurus) need an organization.

The organization of your metadata will always be compliant with the XML standard. On top of XML several other, more specific standards have been created to organize your metadata.

Chapter 5 explains how to chose the type of metadata that fits your needs.

User interface design

We can distinguish many different steps a reader goes through when searching for information. Each of these steps can be supported in one or more ways. Chapter 6 discusses which steps can be distinguished and how these steps can be supported by the user interface

Information Retrieval Test

Of course you want to show that all your work has created added value for the different readers that you distinguished in the stakeholder analysis. You want to show that you are meeting the different requirements that were decided upon in the Information Retrieval Analysis. You could even compare a situation without metadata to a situation making use of metadata. Chapter 7 explains how to

- set up,
- conduct,
- record data and
- analyze whether or not you meet the information retrieval requirements

Chapter 2: choosing the structure of your metadata

Overview

Roughly there are the following types of metadata:

Name	Description
RDF	RDF contains the following information: • The location of the information • The name (title) of the information • The source of the information
Topic Maps	Consisting of • Topic • Relation with other topic • Index Each of these have a class and an instance of a class. So Topic Maps knows 6 different types of metadata
XFML	It is hard to find an interface that takes advantage of all the information you can store in Topic Maps. Therefore a subset of Topic Maps was created: XFML. With XFML you can create a faceted interface. Faceted interfaces are explained in chapter 6. XFML contains • Facet (that can contain other facets), similar to Topic Class • Value similar to Topic Instance
OWL	

Choice

The most simple type of metadata is RDF. It is nothing more (and nothing less) than adding keywords to your pieces of information.

I recommend choosing XFML because it is attached to a very effective interface (see chapter 6). A shortcoming of the current XFML standard (1.0) is that it allows only one Value on one Facet for every piece of information. The current standard does not allow that a piece of information has 2 or more values on the same facet.

Topic Maps are a very powerful way of describing your information. Like Object Oriented Programming in the beginning, we are not sure yet how to turn this power into added value. Therefore I recommend against going to Topic Maps today.

Chapter 3:

Creating a user interface

Steps

When searching, the following needs to happen (by the machine or the human):

Step	Reason
Formulate the question into words	Chose the words, put the words into an order
Correct typo's	A typo may prevent the right information to be found
Translate words into synonyms	Because most search engines work with strings, the reader must make sure he uses all the strings with the same meaning
Translate words into other languages	If you want to search in all languages
Apply a search algorithm to the domain	
Display search result, most relevant search result first	
Get an overview of the search result	
Browse through the search result	
Learn from the search result and decide what may improve the search	

Aids

Machines can do some of the steps above. Also the reader can do some of these tasks himself. When you decide the reader should perform a task, make sure you offer support for performing the task in your user interface.

On the Internet there are lots of tools available that help you spell, translate words and find synonyms.

What is important is that you show the reader that all the steps above help him to perform a search. This means for your interface design that the steps above are visible when appropriate and the reader can perform them easily.

Browse or type

Searching is always a combination of typing in something and choosing from these search results. The reader can only focus on one activity at a time,

- either thinking about the words he needs to type in or
- choose from the choices he gets offered

A reader cannot do both at the same time. There is not something in the middle between thinking of new words and making choices at the same time.

Improving Browsing

Search results show only a small part of the information found. A reader has to decide which information looks promising and which information should be neglected. This browsing can be improved in different ways:

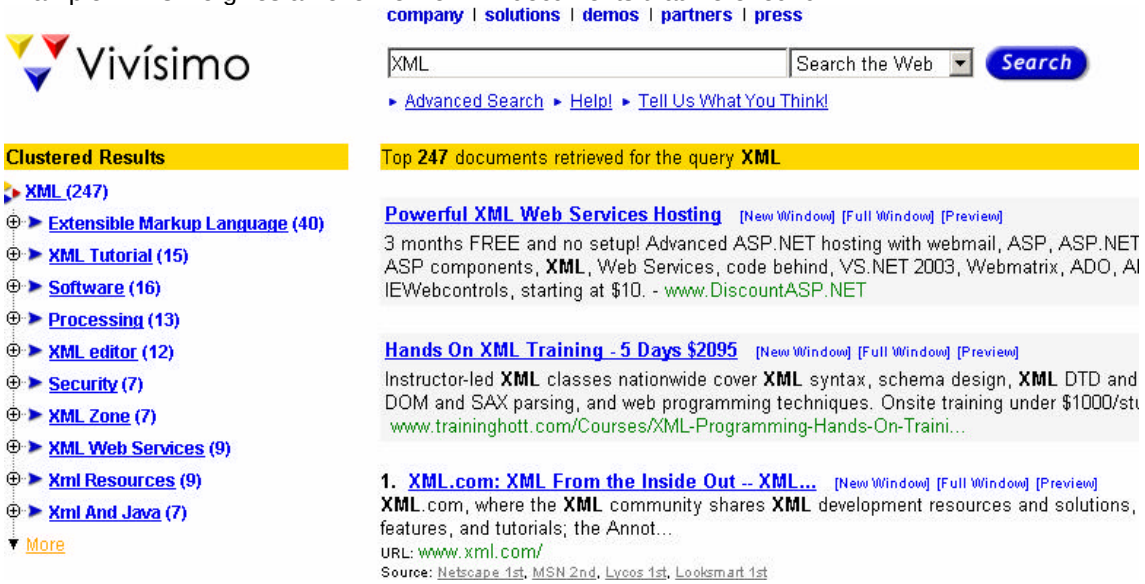
1. Put most relevant search result first
2. Group search results
3. Use a spatial representation allowing the reader to choose a direction
4. Categorize search results on several dimensions, allowing the reader to filter the search result

Putting most Relevant first

This is the preferred technique: do not give the reader 4 million hits, give the one hit the reader needs. So, if your search algorithm can put the most relevant first, stick to it!

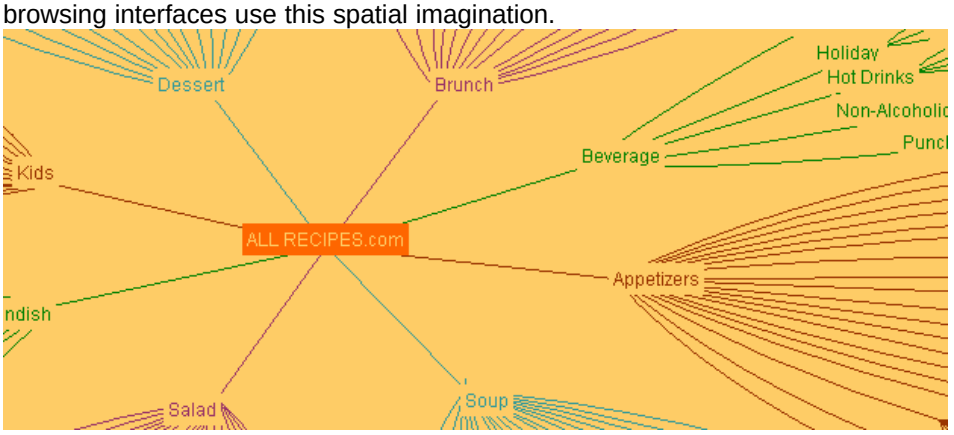
Grouping Search Results

Because a search often returns too many hits, the reader loses the overview. Grouping the search result can help to keep the overview. The search result is grouped according to some terms that the pages have in common.
Example: Vivisimo gives an overview of 247 documents that were found.



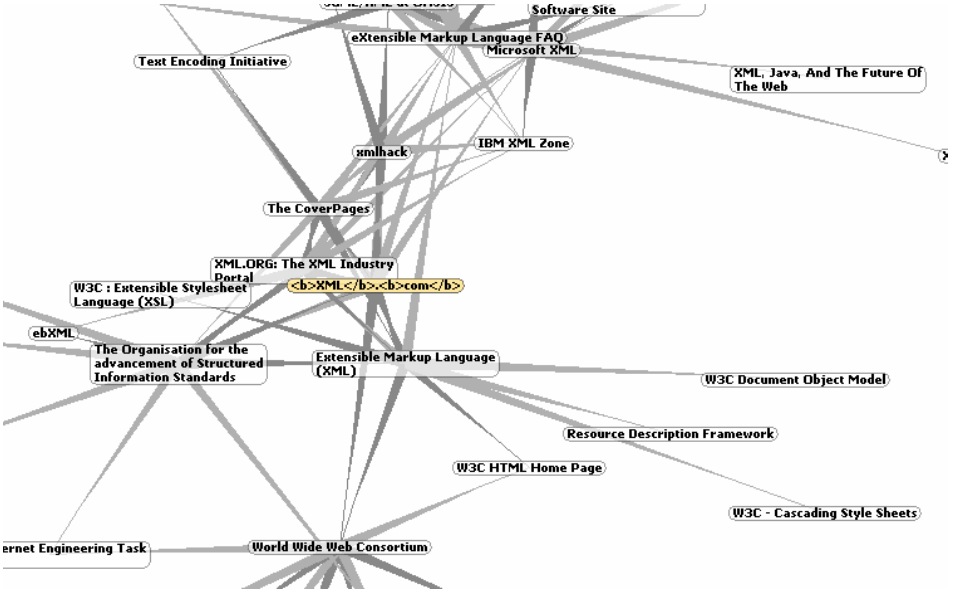
Using a spatial representation

Humans have the ability to imagine that they wander through space. Some browsing interfaces use this spatial imagination.



For example Inxight and the aqua browser allow the reader to choose from several directions. The reader can look ahead and see what the direction will bring. Only the information behind the horizon (formed by the window border) is hidden. By following a direction the reader reveals what lies behind that part of the horizon.

using a spatial representation, continued



Touchgraph goes one step further in using human's spatial imagination. Touchgraph not only allows you to follow directions but when you double click on a location (url) you see the (information)world from the perspective of that page. Touchgraph uses the "similar pages" function from Google to determine what is related. Touchgraph shows you all information related to the page you have double clicked on.

Multi-Dimensional Search

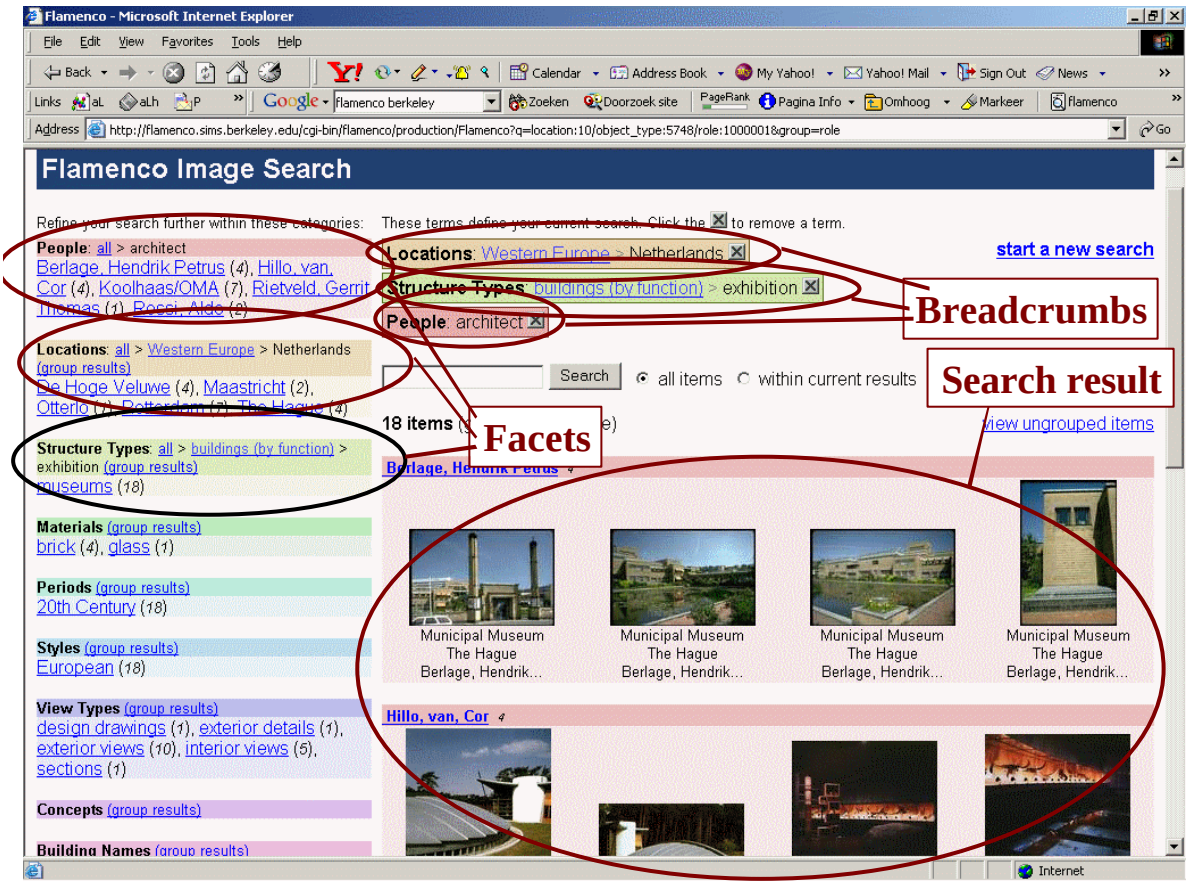
The spatial browsers have the disadvantage of:

1. having three dimensions
2. having a hierarchical data structure

ad 1. Having three dimensions limits the number of directions you can go in to.

ad 2. Hierarchies have the disadvantage that the reader needs to go through a lot of layers before he comes to the desired page. However, the reader always has to start at the top layer. At the top layer it is very hard for the reader to decide which path will take him to the desired page. When finding nothing, the reader needs to go back and start over again. Therefore hierarchies should only be applied when the hierarchy exists also in reality. For example when something is a special instance of something else. Or when something is one of the components of something else.

Faceted interfaces can have numerous dimensions, limited only by what is still effective on a page. Faceted interfaces only have hierarchies when you want them.



The biggest advantage of faceted interfaces is that the reader only makes choices on the facets (dimensions) where he can and wants to make a choice. The reader does not have to make a choice he cannot make.

Every choice will change the visible facets. In this way the reader comes closer and closer to the information he desires. With *breadcrumbs* the reader can:

- Know where he is, what he has chosen
- Undo one of the wrong choices without removing all other choices

Chapter 4

Information Retrieval Test

Orientation

Now it is time to harvest. During the Information Retrieval Test you can see whether readers can easily find the information they need. You can tell whether you have chosen the right keywords, synonyms and user interface. During the Information Retrieval Test, you test the requirements you decided upon during the Information Retrieval Analysis:

- Ratio:

$$\frac{\text{successful searches}}{\text{search attempts}}$$

- Number of search refinements before finding
- Time before finding
- Precision: how much non-applicable information does the reader find?
- Recall: how often is information not found although it is present?
- How applicable is the information?

Setup

You can setup the Information Retrieval Test by:

1. Select a sample of the tasks you found during the Stakeholder Analysis
2. For every task, think very good about the instruction you will give:
   - What is the goal?
   - How does the reader know he has succeeded?
   - What things can the reader use?
   - Is time or any other criteria important?

   It can be good to try out the instructions before you run the actual test
3. Select users that belong to the target audience. There are not shortcuts here. You cannot draw conclusions based on readers that do not belong to the target audience.
4. Create an environment that resembles the environment in which the system will be used

Recordings

During the test you want to know what your reader experiences. Therefore it is good to let the reader think out loud. It is good practice to make both

- Audio- and video recordings, next to a
- Logging of the computer session

Software exists that allows you to record all three sources. Important is that this software keeps the synchronization between the different sources.

Instructions

Your instructions to the subject of experiment should contain the following:

- "Welcome to the test of <the name of your system>"
- "We have developed a new (version of) the <system name> and we want to see whether it is working better in finding the information that you need. So we are not determining whether you do well, we want to know whether the system does well. Therefore, the results will be anonymous, your name will not appear in our findings."
- "To understand what happens during the test you would help us a lot when you could say what you are thinking. Therefore we want you to think out loud. In order to practice think out loud, we will first do a different task before we start the test"
- "Here is the first task. Please read the instructions carefully."
- "Are the instructions clear? If yes, please go ahead and try to achieve the goal described in the task."

Record

While the face and the voice of the reader is recorded on video, the screen is recorded on a log file, you sit on some distance (possibly behind a one-way mirror) and write down everything that strikes you. Don't think too long about what to write. Just write everything that you notice. After a couple of observations you will get better in paying attention to the more 'important' things. Never intervene. The only moment when you may want to intervene is when the reader stops talking about what he thinks.

After the reader has accomplished the task (or gave up), ask the reader for reactions: Was it hard? What was unclear? What method did he follow? Did he enjoy it? What would he recommend?

Analysis and conclusion

Do not focus on the numerical data only. Observations, the reader's opinion, your impressions, all are important. Therefore go through several iterations of summarizing / concluding on what you have found. A part of your conclusions can contain numerical information but another part of your conclusions can be descriptive.