# Building product information search by images

Jean-Claude Bignon[1], Gilles Halin[2], Walaiporn Nakapan[3]

[1]Professor at Nancy School of Architecture
[3]Ph.D. student at National Polytechnic Institute of Lorraine
CRAI (Research Centre in Architecture and Engineering) UMR Map, N°694 CNRS
Nancy School of Architecture
2, rue Bastien Lepage
54000 Nancy
France

[2]Professor at University of Nancy 2
Pôle Lorrain de Gestion,
U.F.R. Maths Info.
13, rue Michel Ney C.O.75
54037 NANCY Cedex
France

## ABSTRACT

Building product information is required during the architectural design and technical design. The common access to the technical information system is the multi-criteria search mode. This search mode is adapted to the situation where an architect has a precise demand of information. But most of the time, the architect looks for ideas and wants to obtain many illustrations of product uses. Therefore, the system has to propose another search mode adapted to the situation where the demand is still fuzzy. Considering that the architect has the capacity to think with image and that an image can generate easily ideas, then a search by images seems to be suitable to the situation where an architect looks for ideas. The web is an inexhaustible resource of images we can exploit to provision an image database on a specific area. The system we propose allows making building product information search with images extracted from the web. This article presents the method used to extract images from web sites of French building product companies and how these images are used in an interactive and progressive image retrieval process.

## 1 INTRODUCTION

The support of technical information usually comes in forms of paper catalogue, CD-ROM, or web site. In order to facilitate the access to the building product database, the HYPERCAT project (Halin et al. 1999b) presents a data organisation of the technical information relatives to the building products and materials. This data organization is use in a building product search service on the web (http://www.crit.archi.fr). At this search service, a multi-criteria search mode (by product type, constructive function, materials, etc.) is proposed to the professionals in building industry (architects, constructors, engineers, etc.).

In this context, we consider that the classic multi-criteria search has a good efficiency but it is not the most relevant data access mode in our particular domain: the architectural design.

In a precise technical information research, the user wants well-known references (name of a product, a brand name, etc.) so he can easily formulate his demand. The existing multi-criteria have proved its effectiveness of search for this kind of situation where the architect knows exactly what building product he is looking for.

But in numerous design situations, the need of technical information cannot be precisely formulated. The architect looks for general ideas and he wants to know which kind of products is available on the market before beginning the design. To response at this fuzzy technical information demand, it seems worthwhile to use adapted search modes.

In our society of visual representation, image can be a new support for technical information research. Therefore, the construction of the technical information search by images would be able to complement the classic multi-criteria search. Our works aim to set up methods and tools for image retrieval. They use the best potentialities of this media and the capacities of the architects to reason with visual representations.

In this article, we present in first part, how the image can be used as a support in a technical information research. The second part proposes a solution to extract and to index images from the web to provision an image database on building product. In the third part, the research process based on the image interactivity is described. Finally, the last part presents the two applications, which implements our propositions.

## 2 USING IMAGE AS A SUPPORT FOR TECHNICAL INFORMATION RESEARCH

The use of image as a support for technical information research presents advantages and unexpected consequences.

### 2.1 Advantages

The reasons of the efficiency of the image are numerous. We shall present some of it.

#### *2.1.1 Neuro-Physiological Reasons*

The most recent studies let us know that 1.5 million of "sensitive fibers" assure the transfer of the sensory information (sight, smell, hearing…). One million of them are dedicated to the visual perception. This observation shows that the visual representation is dominant in the sensitive representation of the brain.

#### *2.1.2 Informative Reasons.*

The image contains a large amount of information interpretable in a short time. This characteristic differentiates it from the other data support. Textual or oral information requires certain linearity of reading or of hearing, in order to understand it. This characteristic is due to their sequential semantic construction. On the contrary, image delivers an immediate global message. The whole semantic construction requires synchronous mechanisms. The morphology of the image's subject, its topology, its colours, the luminous intensity are so many phenomena that are perceptible at the same moment. This particularity of image contributes to the direct transfer and assimilation of information.

*2.1.3 Cultural Reasons.*
Although the oral and the writing remain long-lived, we are in a culture which gives more and more place to the image. Various reasons are identifiable: some are very old as the proof effect, which plays the visual perception in the occidental culture (I believe what I see). The others are more recent such as the search for increasing speed in the information exchange. We will also evoke the technical mastery acquired today in the image (photography, cinema, television,), which contributes to give to this media a dominant role in our information society.

*2.1.4 Professional Reasons.*
Among the architects, the image (drawing, photo…) plays an important role in the design mechanisms. Image is not only a raw material of the creation, but also a way to see and to perceive a problem. These heuristic functions of the representation are recognised (Boudon et al. 1988). But this media assures also a patrimonial function. It represents the support for the constitution and the transmission of doctrines and spaces of architectural references. Models are transmitted in image form. This visual culture leads the architects to develop a specific intelligence, the "visual/spatial intelligence" (Gardner 1992) in which a number of mechanism of reasoning are built "by the image". Thus, image plays a major role in architect's way of reasoning, which is essentially abductive.

### 2.2 Disadvantages

If image holds numerous advantages, it has inconveniences, which can produce informative disorientation or misunderstandings.

*2.2.1 Polysemy*
The semantic richness of the image has a consequence in the plurality of readings and interpretations. The same image can communicate very different information, which depends on the point of view adopted by the receiver. The code of the standardised representations (such as technical drawings) is common to the transmitter and to the receiver. The message is identical if the transmitter and the receiver look at the image from a common space of references. But in most of the images, especially photographs, a same information can refer to different codes of reading, which is the source of misunderstandings.

*2.2.2 Information Overload*
This situation occurs when an image represents several different objects. A caricatured example is a composite image, which consists of several smaller images (a mosaic image). But the situation of multiplicity of objects is also found in images where the scene contains many objects. The reproduction of the information disturbs a precise identification of the object that the transmitter wants to show.

*2.2.3 Information Deficiency*
Particular stylistic forms needs more efforts of interpretation. Moreover, such stylistic image can mislead the informative contents of the image. There are so many of such situations, like the metonymy or the metaphor. But there are also the technical distortions as the images are cut or shaded off gradually (weak resolution, limited number of colours, size reduced from original, etc.). This rich quality of the image can turn out to be penalising for the situation where one needs a clear interpretation.

*2.2.4 Semio-Graphic Ambiguity*
Some images are reworked to improve their attractive function, such as graphic efficiency, aesthetic quality, or fashion effects. This graphic work produces a transition from the immediate representation universe to the more allegorical cultural universes. The interpretation of an image is easy if the image is similar to the real world object. But the graphic work makes this similarity more distant. The overall representation becomes more important than the represented object. Therefore, the interpretation of represented object is troubled. This graphic ambiguity passes on values that formal similarity cannot.

## 2.3 Question of a Relevant Image

The inconveniences of image presented above lead to the following question: *"What is a relevant image for an architect when he looks for a product or a building material?"* There are three categories of relevance:

- visual relevance,
- semiological relevance,
- graphical relevance.

*2.3.1 Visual Relevance*
There has to be a strong visual similarity between the represented object and the representative image. One will notice that this resemblance takes two conditions: the represented object must be known and its representation must be made in a similar universe of interpretation.

The visual similarity is an effect producer. The theory of efficiency can be applied to the visual relevance. An image is relevant if it produces enough effects to balance the efforts needed for its interpretation (Sperber and Wilson 1989).

Here are some criteria that favour this visual relevance:

- The more an image is similar to the real world object, the more the image is relevant,
- The more an image produces effects, the more the image is relevant.

*2.3.2 Semiological Relevance*

Beyond the semiological decoding, the interpretation of an image implies inferential processes. These mechanisms are based on the information that is not coded in the image, generally called context (Reboul and Moescler 1998).

This notion of context means the immediate context of the image (the text that is attached to the image, the other images that is in relation with the image). But the context also means the reader's context of interpretation. The semantic aptness of an image depends on the meaning of the image "in itself". It depends on the relation that exists between this image and its appropriate context. It also depends on the relationship between the reference universes of the interpreter (Boullier 1999).

The images are represented in a product catalogue, or a web site, etc. The extraction of images for database provision means removing them from their context. This isolation makes removed images insignificant and causes misunderstandings. A misunderstanding appears when the transmitter produces an image with information that the receiver doesn't have. For example, the information does not exist in the image itself, or the information does not exist in the reader's universe of interpretation (Reboul and Moescler 1998).

In such approach, the relevance of an image can be defined from two criteria:

1. The more an image requires an interpretation effort, the less the image is relevant,
2. The more a context of an image is interesting, the more the image is relevant.

*2.3.3 Graphical Relevance*

The graphical relevance plays an essential role in the selection of relevance image. The form of an image (width, height, and proportion) can verify the graphical relevance in a certain limit, without knowing its content.

A too big image can be composed of smaller images, which complicates the interpretation. A too small image can be an icon or serve as decorative element. Concerning the proportion, an image with a width three times its height can be a banner of a Web page, far from being a photograph. The problem is how to identify if an image is too big, or too small, or not in a right proportion.

This approach lead us to a conclusion that an image is relevant graphically from the following criteria:

- The more an image's size is in a limited interval, the more an image is relevant,
- The more an image's proportion is in a limited interval, the more an image is relevant.

*2.3.3 Conclusion*

For the selection of relevant image, there are three important principles to apply*:*

- *a principle of visual analogy,*
- *a principle of legend,*
- *a principle of form.*

*2.3.3.1 The principle of visual analogy*

The principle of visual analogy is applied in order to select images that produce most effects. The relevance of an image depends on the following criteria:

· there has to be a similarity of colour between the image and the usually dominant colours of the represented product,
· the object must be represented entirely. The more it is cut, the more it is necessary to interpret the missing parts,
· the scale of representation has to allow the represented object to occupy an important surface area on the image. The less important the surface area, the more of the other image can play the role of main subject,
· it is necessary to maintain elements of an object's usual environment in the object representation (example: a type of faucet will be better noticed if it is situated near a kitchen sink or near a boiler).

*2.3.3.2 The principle of legend*

The principle of legend is applied to minimise the effort of interpretation. Keywords are associates to an image in order to refer its universe of interpretation. These criteria are valid only for our particular approach, in which we give greater importance to the recognition effect than to the suggestion effect. The relevance of an image depends on the following criteria:

· an image selected should have a nearby context or a legend,
· the context of the image is interesting if the keywords found correspond to terms in the thesaurus.

*2.3.3.3 The principle of form*

The principle of form helps selecting an image without knowing its content. It eliminates the image that causes interference. The criteria are as follow:

· an image's size (width, height) should be in a limited interval, corresponding to the size of a photograph,
· an image's proportion (width/height) should be in a limited interval, close to the proportion of a square.

These principles allow the selection of pertinent images. They should be applied to the image extraction and indexing process. The principle of form can be applied to an automatic process of image extraction. The principle of legend can be viewed as possibly automated if the image's legend is a text, not an image of a text. However, the principle of visual analogy still needs a human intervention.

However, web images are chosen as domain of interest. The following paragraph explains the method of image extraction and indexing from the web.

## 3. IMAGE EXTRACTION AND INDEXING FROM THE WEB

The image from the web comes to our interest because they are numerous, accessible, and in various form. They can be extracted automatically applying the principle of form and the principle of legend above. This automatic extraction facilitates a provision of the image database.

### 3.1 Image extraction problem

An image in a paper catalogue is usually accompanied by text. So is the image from a building product web page. In a web page, an image and a text are found next to one another. Therefore, an image extraction and image indexing using its context can envisage an automatic process.

The problematic of the image extraction is how to select relevant images from a web page. An image extraction from the web is very costly if the system has to verify every image in every page and follow every link in a page. The extraction rules should help organising the order of verification. If the system can avoid the redundancy of task, then the cost can be reduced.

Another problem is how to index relevant images. Our image retrieval process is a context-based one. Once an image is extracted from a web page, it needs to be indexed with terms in its context. However, a retrieval system performs better if a document is indexed with a limited number of terms (Leloup 1997). Therefore, we prefer that an image is indexed with terms in the thesaurus. The problem is how can the system associate the context of an image to terms in the thesaurus.

### 3.2 The Extraction Method

The method consists in applying successively two processes:

- an extraction process guided by a decision tree,
- an indexing process using the thesaurus terms.

#### *3.2.1 An Extraction Process Guided By A Decision Tree*

A decision tree is a good representation of the extraction process because of its sequential property. Each node of the tree represents a question. At a node, the system chooses a branch of the tree, which represents the answer. The chosen branch leads to another node. Therefore, a decision tree forms a series of questions that the system has to answer.

This form of representation is served as a guideline for a relevant image extraction. The selected image comes to the end node (six) with the branch “yes”. Each node of the decision tree (Fig. 1) can be described in the following way:

1. “The Web page is in a limited distance?”

The distance of the Web page is set to zero by default, which means that the URL stays in the same domain (parent URL). If the distance is set to 1, the system will accept the pages that are referred to by a page that has the same parent URL.

2. “The page is written in French?”

The choice of language used in the page eliminates all the pages that are written in foreign languages (other than French). Because images found on that page are impossible to be indexed by the words of the thesaurus in French language.

3. “The image is in an interesting page?”

Interesting images seem to be found in appropriate pages, which means the product presentation page. In such way, when the extraction process looks for an image of building product, the image and its context are usually relevant. The list of forbidden words in <A HREF> tells the extraction process where the page doesn’t worth to be investigated (presentation, history, address, contact, link, etc…).

4. “The form of image is OK?”

The principle of form is applied here. These criteria examine its graphical relevance of the image considering its width, height, and proportion (width/height). They help eliminating the decorative images commonly found on a Web page. The criteria are validated as follows:

- 60 pixels ≤ width ≤ 610 pixels,
- 60 pixels ≤ height ≤ 60 pixels,
- 0.58 ≤ proportion ≤ 2.1.

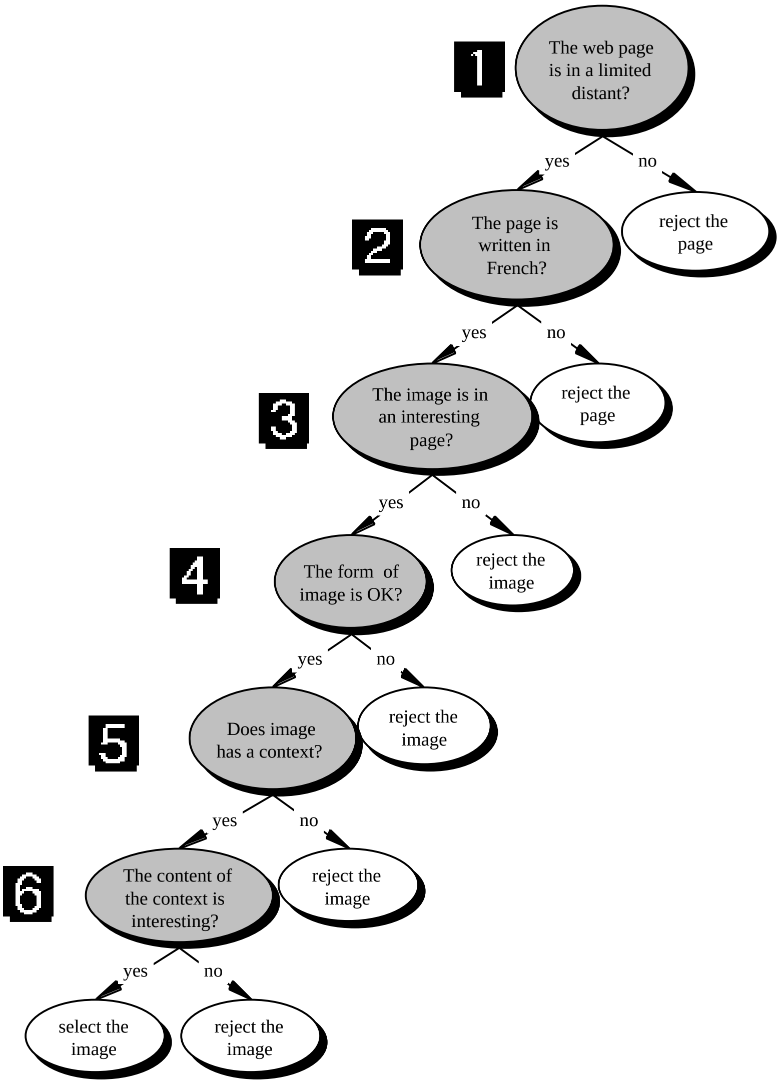
5. “Does image has a context?”

The principle of legend applied here. The criterion is that an image selected should have a nearby context or legend. The image is rejected at this stage even if there is not a context. Without a context, the content of an image is unidentifiable. However, a context of an image can be a context of another. The calculation of the distance between each image and each context helps determining which context belongs to which image.

6. “The content of the context is interesting?”

An another criterion of principle of legend is applied at this stage. The extraction process selects an interesting context for image indexing. The content of the context is interesting if we find that it is described by words that has equivalent in the thesaurus.

**Figure 1: Decision tree**

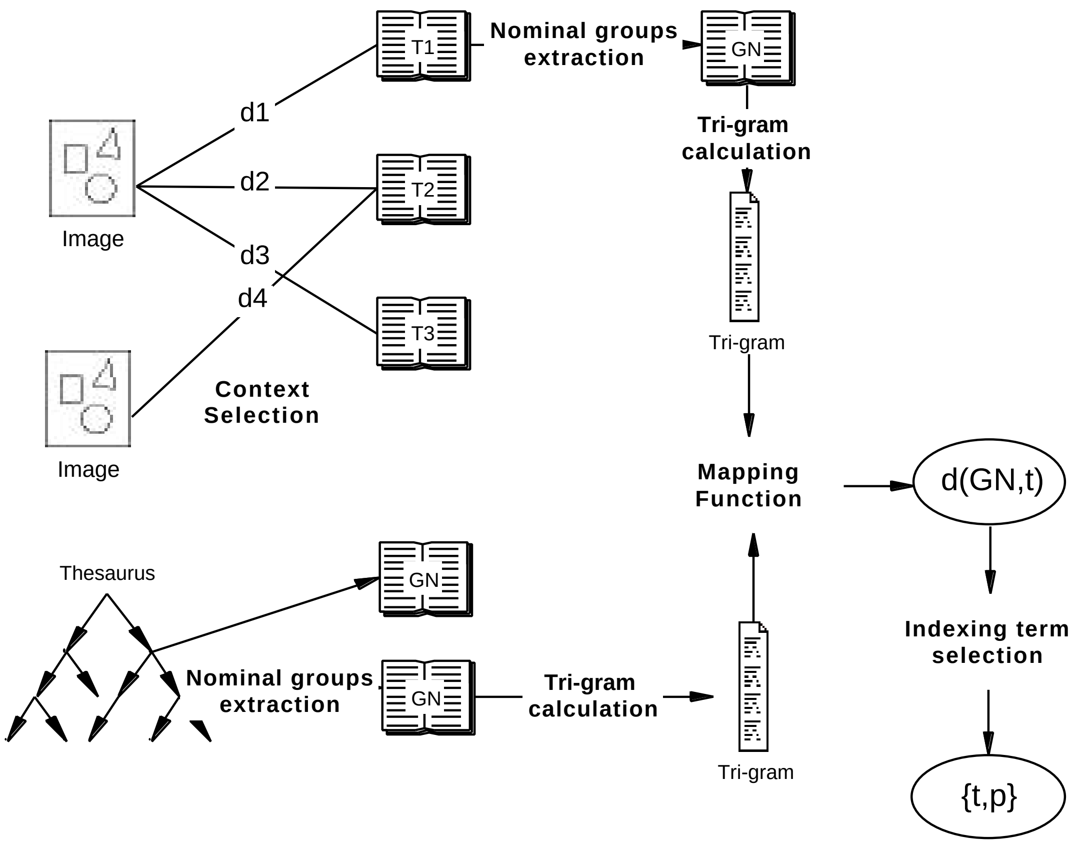



The process extracts the images that pass through the branch “yes” of the last node (question 6). The result is a list of image and the keyword found in their contexts. Then these keywords are mapped to terms in the thesaurus. This form of indexing is a controlled text indexing using the thesaurus.

*3.2.2 An Indexing Process Using the Thesaurus Term*

In order to fight against noise and silence in the image retrieval process, the controlled text is used as indexing vocabulary. Therefore, keywords in the context will be mapped with the thesaurus terms.

**Fig. 2: The indexing process**




The indexing process (Fig. 2) can be described as followed:

On one side, the process transforms each thesaurus term into nominal groups and calculates the tri-gram. This process is to be done only once, unless there is a modification of the thesaurus

On the other side, the process selects the context (T) of each image. Then, the nominal groups (GN) are extracted from the context and then transformed into tri-gram (TG).

Next, the mapping function calculates the distance (d(t,p)) between each context's tri-gram and each thesaurus's tri-gram. Then the process ranks terms by weight for each indexing domain. The term with the highest weight of each indexing domain is selected as indexing term of the image.

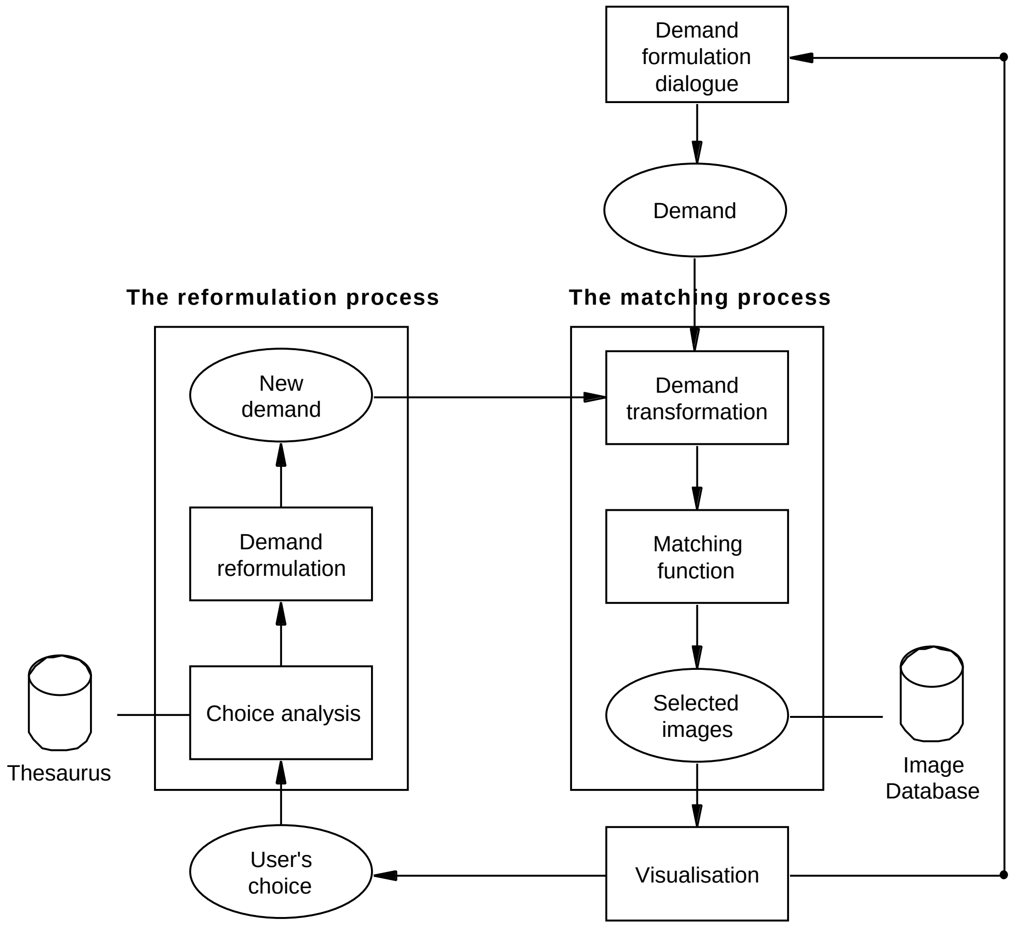
The indexing terms are information that helps retrieving an image. They are used in an interactive and progressive image retrieval. This search mode uses the relevance feedback (van Rijsbergen 1979) and the vector space model as a mapping function (Halin et al. 1990).

## 4 AN INTERACTIVE AND PROGRESSIVE IMAGE RETRIEVAL

Understanding the user’s need is very important in technical information search. However, the system cannot give what the user finds acceptable if it doesn’t ask what he think of the result (Eakins 1996). A dialogue between the user and the system can help to better understand the user’s need.

Consequently, an interactive and progressive image retrieval is proposed. This form of context-based image retrieval better uses the interactivity of the image (Halin 1989). The technique of the “Relevance feedback” makes possible for the system to understand the user's need more precisely.

**Fig. 3: The process of interactive and progressive image retrieval**




The principle of the search process (Fig. 3) is as following:

- the user formulates a simple demand,
- the system analyses the demand and proposes a first set of images to the user,
- the user chooses or rejects images. He can also stay indifferent on an image,

- the system analyses these choices and reformulates a request in order to select new images that are proposed to the user.

The process stops when the user obtained a set of images, which he considers relevant. He can eventually ask for technical information that the images represent.

5 THE APPLICATION

Two applications have been realised. The first application is for image extraction and indexing from the web. This application provides images for the second application, which is an image search engine for technical information search.

**5.1 Wimex-Bot**

The Wimex-Bot is a robot for image extraction from the web. It is specialising in building product images. Written in Java, it extracts and indexes building product images on web pages using their HTML source code.

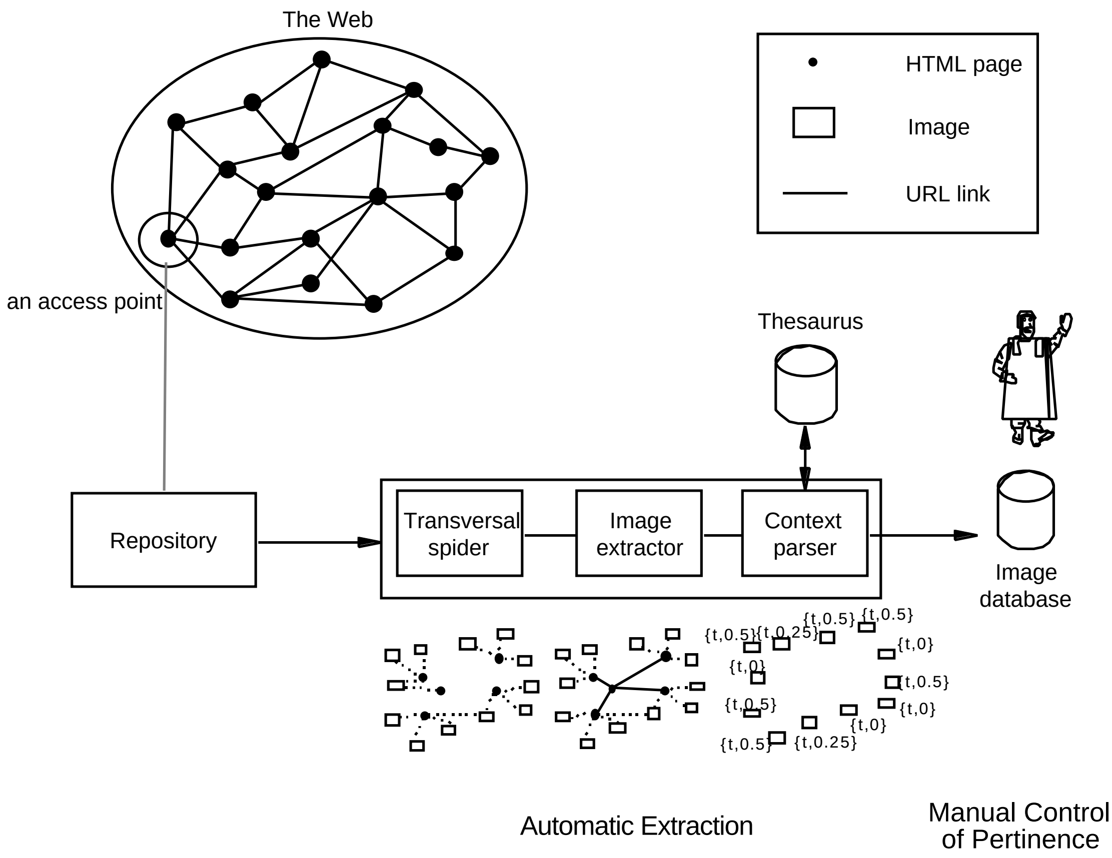
Related work is the *Marie project* of US Naval Postgraduate School (Rowe and Frew 1997). It also extracts and indexes the image using the HTML source code. However, the system finds the caption of the image in stead of its nearby context. Another early work in image and text extraction on the web is the *Image Excavator* from School of Computing Science, Simon Fraser University, Canada (Zaïane et al. 1998). This web agent uses image textual information, like HTML tags in web pages, to derive keywords in English language. The agent traverses on-line directory structures like Yahoo!. Consequently, it can create hierarchies of keywords mapped to the terms on the directories in which the image was found.

The Wimex-Bot is composed of two principle parts:

- the automatic extraction and indexing,
- the manual control of pertinence.

These two principal parts make the application a semi-automatic one. The principle of Wimex-Bot (fig. 4) can be described in the following way:

**Figure 4: The principle of Wimex-Bot**




The automatic extraction and indexing is composed of three robots. They are inspired by three spiders of the image and video collection process for WebSEEk (Smith and Chang 1996). These robots begin with a repository and an access point to the web. A list of web page is selected from the building product database. The “Transversal Spider” travels the web in order to find every HTML pages linked from the beginning page. The distance from the origin page delimits the page to investigate.

However, only interesting pages are selected for image extraction. The forbidden words (address, contact, history,…) are listed in order to tell the “Image Extractor” which page doesn’t worth to be investigated. The potential pages of each site are selected. Then, the “Image Parser” verifies images on each page if they correspond to the extraction rules or not. The first filter by form (height, width, and proportion) is applied here.

Then the “Context Parser” verifies if an image has a context around it. However, it validates only interesting context. The context is interesting when there are thesaurus terms in the context. At this stage, the indexing process indexes the image with thesaurus terms using its context.

At the end, an intervention of administrator seems necessary in order to control the pertinence of extracted data. He validates the pertinent images and indexing terms. These data are then stored in the database.

## 5.2 BATIMAGE

BATIMAGE (Halin et al. 1999a) is an application for image retrieval. This prototype was produced by using a sample of the images extracted from the Web sites of French building products. This prototype uses the interactive and progressive image retrieval. This process can be described as follow:

- a first window request the user to choose a topic, the type and the number of the images he wishes to visualize (construction the first request),
- this request allows the selection of a first set of images. These images will be displayed in the shape of a mosaic (cf. Fig. 5),
- the user can choose or reject images. He can also not give an opinion on certain images,
- this choice carried out, the system analyze the choice and send a new request to BATIMAGE,
- the new images are selected and presented.

**Figure 5: BATIMAGE**

The process ends when the user asks for a list of building products that his choice of images represent. He can asks for further information about the building product provider, etc.

## 6 CONCLUSION

The HYPERCAT project presents an organisation of building products and technical information for experts in the building industry. However, the classic multi-criteria search is not the best search mode in our particular domain of architectural design. Since the architects can reason better with the visual representation, image seems to be an ideal support.

The advantage and inconvenience of image as support for technical information search are presented. They lead to an analysis of pertinent images for this activity. The conclusion of this analysis is that three principles have to be applied for pertinent image selection. Some principle can be automated and some need human intervention.

The inexhaustible Web resource is chosen for the provision of image database. A method presented in order to extract web images and to index them. First, a decision tree is proposed as a guideline for image extraction process. Second, the indexing process is a controlled text indexing. It indexes images from the web page using thesaurus terms.

Then, the principle of the interactive and progressive image retrieval is proposed in order to understand the user's need. It is applied in a conception assistance tool of technical information search by images. This application gives the user ideas of products available and helps him to formulate his demand of information. The result of his search is the technical information of the building products.

The perspective of this research is the improvement of image extraction and indexing from the web, which still needs human intervention. Apart from that, the image database needs a regular automatic update.

## 7 REFERENCE

Boudon, Ph., Pousin F. (1988), *Figure de la conception architecturale. Manuel de figuration graphique*, Dunod, Paris.

Boullier, D. (1999) *l'urbanité numérique. Essai sur la troisième ville en 2100*. L'Harmattan, Montréal.

Eakins, J.P. (1996) Automatic image content retrieval – are we getting anywhere? *Proceeding of ELVIRA: Third International Conference on Electronic Library and Visual Information Research*, Milton Keynes, United Kingdom, April 30 - May 2, 1996.

Gardner, H. (1992) *Multiple Intelligences: The Theory in Practice.* Basic Books, New York.

Halin, G. (1989) Apprentissage pour la recherche interactive et progressive d'images : Processus EXPRIM et Prototype RIVAGE, thèse, spécialité en informatique, Université de Nancy 1, Nancy.

Halin, G., Bignon, J.C., Humbert, P., et al. (1999a): Interactive and progressive image retrieval on the WWW. Application on the building product search. *Presented in Workshop on Multimedia Databases and Image Communication (MDIC'99)*, Salerno, Italy, October 4-5, 1999.

Halin, G., Nakapan, W., Bignon, J.C. (1999b): Navigation par l'image sur le WWW. Le catalogage Hypermédia des produits du bâtiment : le projet Hypercat. *Proceeding of the 5th International Conference H2PTM'99*, Paris, France, September 23-24, 1999, pp. 327-331.

Leloup, C. (1997) *Moteurs d'indexation et de recherche : Environnements client-serveur, Internet et Intranet*. Eyrolles, Paris.

Reboul, H. and Moescler, J. (1998), *La Pragmatique aujourd'hui. Une nouvelle science de la communication*, Seuil, Paris.

Rowe, N. C., Frew, B. (1997): Finding photograph captions multimodally on the World Wide Web. *Proceedings of the AAAI Spring Symposium on Intelligent Integration and Use of Text, Image, Video, and Audio Corpora*, Stanford, CA. March 24-26, 1997, pp. 45-51.

Smith, J. R., Chang, S.-F. (1996): Searching for Images and Videos in the World-Wide Web. Technical Report#459-96-25, Department of Electrical Engineering and Center for Image Technology for New Media, Columbia University, New York.

Sperber, D. and Wilson, D. (1989) *La pertinence: communication et cognition*, Ed de minuit, Paris.

van Rijsbergen, C.J. (1979): *Information Retrieval*. 2nd edition. Butterworths London

Zaïane, O.R., Han, J., Li, Z.-N., et al. (1998): Mining Multimedia Data, *Proceedings of CASCON'98 Conference*, Toronto, Canada., Nov. 30-Dec. 3, 1998, pp. 83-96.