Integrating Image Content and its Associated Text in a Web Image Retrieval Agent

Victoria Meza **Jesus Favela**
{mmeza, favela}@cicese.mx
CICESE Research Center
Km. 107 Carretera Tijuana-Ensenada
A. P. 2732. C. P. 22800, Ensenada, B. C., Mexico

Abstract

In this paper we present a search engine that traverse the Web to find images. Our approach combines a wavelet-based content image retrieval technique with a traditional text retrieval method. The two methods are applied independently too build two indices and then combined in a similarity metric which is used for retrieval.

1. Introduction

The explosive growth of the Internet gives us access to considerable amounts of information distributed around the world. The World-Wide-Web, with its graphical interface and the ability to integrate different media and hyperlinks is the main responsible for this growth. Finding useful information through navigation in such a large information space has proved to be difficult (Cheong 1996).

Several tools have been developed in the last couple of years to assist the search of information in WWW. These tools are often referred to as Spiders, since they build indices of information available in WWW while traversing this Web of information. The term Internet Agents is also used to refer to these information retrieval services. However, the term agent is generally associated with autonomous intelligent behavior by the computer science community and most of these Spiders use traditional information retrieval algorithms based on Tf*IDf (term frequency - inverse document frequency) (Salton 1989) and Latent Semantic Indexing (Krovetz & Croft 1992) rather than being knowledge or learning-based. Retrieval engines that do not require an explicit query from the user, but rather infer user needs based on the user's browsing behavior have been proposed more recently (Lieberman 1995; Voigt 1995).

An important limitation of all these systems is that only take into account one type of information available in the Web: Text. Images, audio and video contained in Web pages are ignored by the indexing schemes of these systems. A partial solution to this problem is the use of the text associated to the image's caption to characterize the image (Smeaton & Quigley 1996).

On the other hand, extensive work is being done in content-based image retrieval (Jacobs, Finkeltein, & Salesin 1995; Niblack et al. 1993; Uriostegui 1996). These systems make use of image processing and computer vision techniques to retrieve images from a database using attributes such as color, texture, form, etc.

In this paper we describe a tool, that we are currently developing, to retrieve images in the WWW by combining a content-based image retrieval method based on image decomposition using wavelets with an information retrieval technique applied to the text associated to the image. We perform a parallel search for each technique and the results are combined in a similarity metric which is used for retrieval.

The following section presents the general procedure for Web image retrieval that is being proposed. Section 3 describes how the indices are created and defines the agent navigation strategy. In Section 4 we present the retrieval metric which combines the results of the text and image retrieval methods. Finally, in Section 5 we present our conclusions and describe future work.

2. Web Image Retrieval Process

The image retrieval process is illustrated in Figure 1. The user creates a query in a Web browser by painting an image similar to the one he wants to retrieve using an applet. A Haar wavelet transform is applied to the image (Stollnitz, DeRose, & Salesin 1995). The wavelet coefficients with the highest value and the most important keywords associated with theimage are used to calculate a similarity metric with the images stored in a database. The user then obtains a list of pages with ranking images.

Several issues need to be considered in the implementation of the system:

How will the indices of the images in the database be built?

The issues to be addressed here are:

a) The Web navigation strategy to build the image database.

b) The indexing mechanisms for image attributes and related keywords.

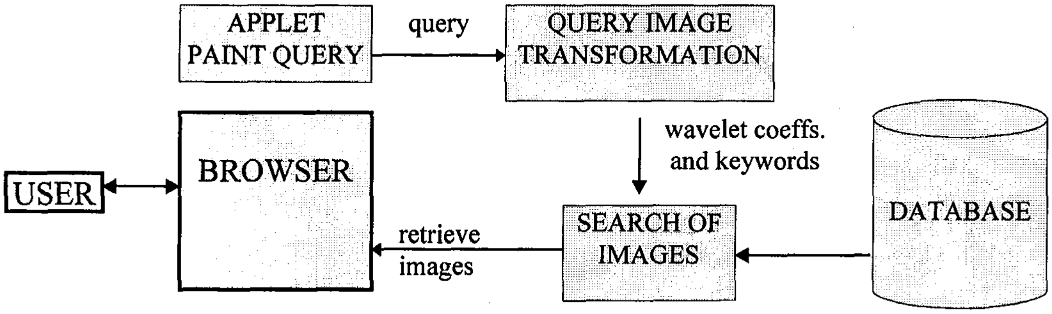


Figure 1. Web image retrieval process

How will the images be retrieved?

This issue can be divided in two parallel searches. The first one being content-based image retrieval and the second one applied on the text that is associated to the image. The results of these searches are combained in a similarity metric to obtain a single value.

The user interface

The system is being developed using the Java programming language (Sun 1995) and can be used from any Web-browser that's Java enabled. To formulate the graphical query the user is provided with a simple painting system. The user selects colors from a palette and draws on the canvas using the mouse.

3. Building the Image and Text Indices

3.1 Agent Navigation Strategy

An agent is a program with intelligent behavior and delegated authority that provides services for the benefit of a user. A Web-agent normally traverses the Web following hyperlinks from the pages it visits. An agent can be used to search for information, schedule a meeting or make a transaction for the benefit of a user. Agents exhibit autonomous behavior in the sense that it uses internal mechanisms to make decisions as to how it achieves its objectives.

The Internet has provided opportunities for the development of agents that assist in the retrieval and filtering of information. Most of these assistants are search engines that traverse the Web indexing the information found in its nodes. Some of the better known systems are WebCrawler (Pinkerton 1994) and Lycos (Mauldin 1995). WebCrawler uses breadth-first search to do full-text indexing of the titles and content of the documents found using a vector space information retrieval model. Lycos creates an index of the 100 most important words in the document determined by their placement within the text and the frequency in which these words appear in the document with respect to the number of times in which they appear in all the documents.

Our agent is similar to WebCrawler with the characteristic that it will look for GIF images. Once an image is found, the agent will decide what part of the text in the page is associated with the image. In a first approximation it looks for the caption html tag. In the absence of this it determines whether there's text below or to the right of the image. When the page only has an image, the page's title will be used as associated text. Finally, if the image has a mark and there's a link from the mark to the image, the paragraph from where the reference is made will be taken as the text associated with the image.

3.2 How are Image Attributes Defined?

The system uses two indices. The first one is made from the wavelet coefficients with highest value from the image and the second one stores the most significant words associated with the image. Figure 2 shows the steps taken to build these indices.

The process begins with an html page where an image in GIF format has been found. The image is extracted and resized to 256x256 pixels . To each of the RGB channels of the image we apply the wavelet transform and store the 80 coefficients with highest values as well as its location within the image. These coefficients are stored as the image indices. A similar image will have several of its higher value coefficients in the same spatial location.

As we do with the image, we extract the text that is associated to the image and identify the 10 most important words associated to that image using information retrieval algorithms as described in section 3.2.2.

3.2.1 Wavelet Description of Image Content

Wavelets are a mathematical tool used for function decomposition (Daubechies 1988). They offer an elegant technique for representing and analyzing signals in a multiple levels of resolutions. Wavelets have their roots in approximation theory and signal processing and have been applied to image compression, computer graphics, and other fields (Stollnitz 1995).

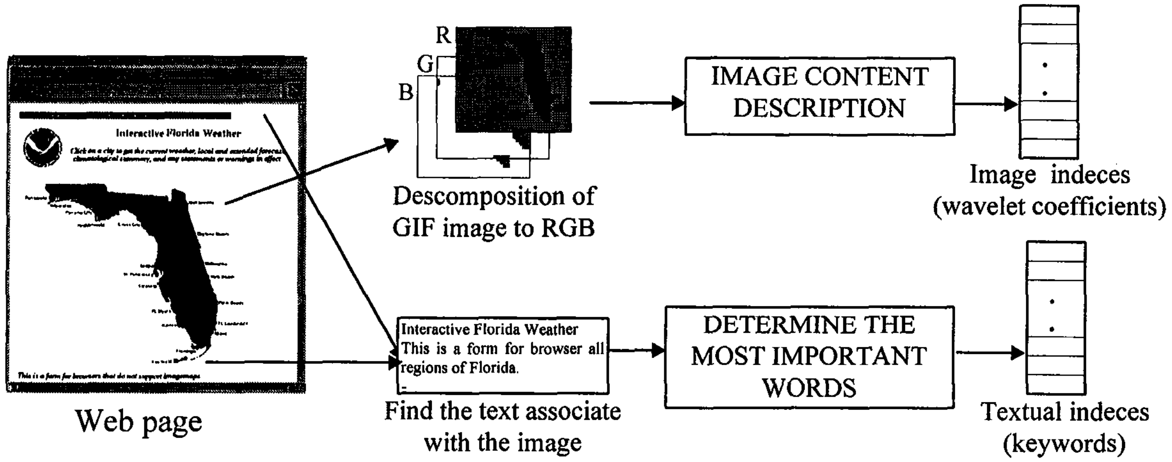


Figure 2. Determining the image indices.

In our application we use the simplest of wavelets: the Haar wavelet basis. This wavelet transform can be estimated quickly and tends to produce blocks in the details of the image, which is convenient for our application given that the query image, drawn by the user, is made of regions constant color.

The wavelet decomposition of the image is applied in two dimensions, first to the rows and then to the columns (Stollintz 1995).

Once the image is transformed we obtain a matrix of wavelet coefficients $W_{256x256}$, which can be used to obtain the original image. Of these coefficients we identify 80 which have the highest absolute value and keep only those as a signature of the image. We store only the location of these 80 coefficients for each of the RGB components of the image.

The similarity metric used for comparing images from their wavelet coefficients is taken from (Jacobs, Finkeltein, & Salesin 1995):

$$||Q,W|| = W_{00} \, |Q[0,0] - W[0,0]| + \Sigma \, w_{i,j} \, |Q[i,j] - W[i,j]|$$

where Q[i,j] represents the query image's wavelet coefficients, W[i,j] the coefficients of the image with which we are comparing the query, and $w_{i,j}$ are weights. This metric is applied to each of the RGB components of the image.

3.2.2 Textual Descriptors of an Image

To identify the text that is associated to the image we have identified the following criteria from the observation of a number of html pages:

1. If an image has a caption html tag, its text is associated with the image.
2. If there is text to the side of the image, this text is considered to be related to the image.
3. If there's a link in a paragraph to the image, the whole paragraph is selected
4. If the page only has a title and the image, the title is the associated text.
5. If the page only has an image, the title of the image is the associated text.

Once the text has been identified we proceed to determine the 10 more important words within the text. These are the words that better discriminate the document with respect to others. To do this, we use traditional information retrieval techniques by first removing the "stop words" (those words that being so common are very poor discriminators). The system then uses the Tf*IDf (term frequency - inverse document frequency) algorithm to assign to each word a weight that gives an indication of the importance of that word in describing the given document. Tf*IDf takes into account the number of times in which the word appears in the document with respect to the number of times in which it appears in the database of documents.

For each image we store the 10 most important words and its associated text. A dot vector between the terms and weights found in a document (d_i) and those found in the query (q) are used as similarity metric to determine weather a document matches a given query:

$$s_t = \langle q, d_i \rangle$$

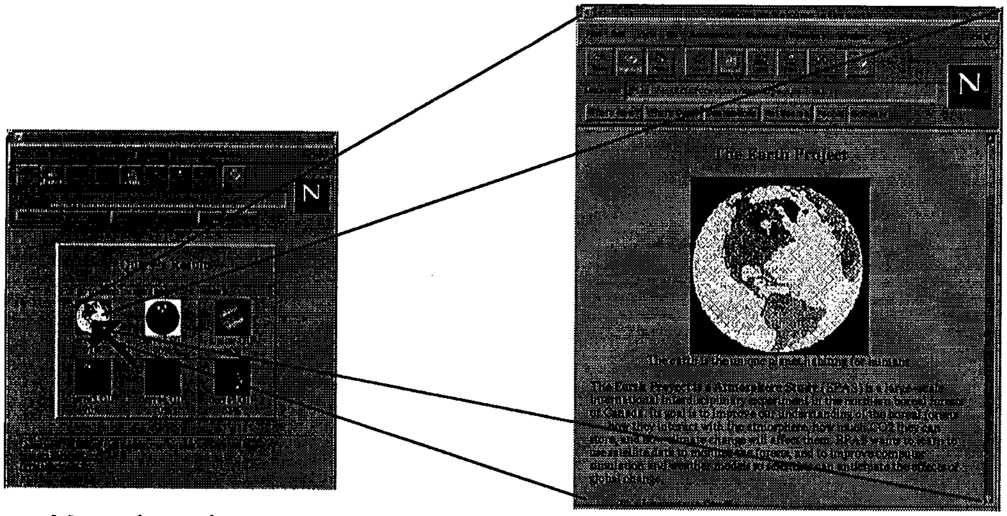
4. Relevance Metric

The relevance of given image query with respect to an image in the Web is determined by performing two parallel searches. One based on image content, and a second one based on the keywords that represent the text associated with the image. The resulting similarity metric is as follows:

$$R = \alpha \, ||Q,W|| + \beta \, \langle q, d_i \rangle$$

where α and β are weights which could be determined by the user. We plan to conduct experiments to determine appropriate values for these parameters.

Figure 3 shows the display of the results of a query. For



More relevant images

Selected image

Figure 3. The system retrieves the thumbnails of the images that are closer to the query. Clicking on a thumbnail links to the html page were it was originally found.

some of the resulting images the wavelet coefficients made for most of the similarity while others were retrieved mostly from the associated text. We can select one of the images and this will take us to the html page were it was originally found.

5. Conclusions

In this article we have proposed a strategy which combines image and text retrieval techniques for the retrieval of graphical content in the WWW. We have designed an agent that navigates the Web looking for images and determines and retrieves the text that is taught to be associated to the image. From the image and its associated text we build indices applying a wavelet transform to the image and storing the most significant coefficients, and from the text we identify the most significant words using the Tf*IDf information retrieval algorithm.

We are currently implementing the system. Preliminary results of its use are expected by March 1997.

References

Cheong, F. 1996. *Internet Agents: spiders, wanders, brokers, and bots.* Ed. New Riders.

Daubecheis, I. 1989. Orthonormal Bases of Compactly Suported Wavelets. *Comm. on Pure and Applied Mathematics.* 41: 909-996.

Jacobs, C. E.; Finkeltein A.; and Salesin D. H. 1995. Fast Multiresolution Image Querying. Siggraph 95 Conf. Proc., ACM, New York.

Krovetz, R., and Croft, B. 1992. Lexical Ambiguity and Information Retrieval. *ACM Trans. on Information Systems* 10(2):115-141.

Lieberman, H. 1995. Letizia: An Agent that Assist Web Browsing. Workshop on AI Applications in Knowledge Navigation and Retrieval, AAAI-95 Fall Symp Series:97-102.

Maulding, M. 1995. Measuring the Web with Lycos. Third International World Wide Web Conference. http://lycos.cs.cmu.edu/lycos-websize.html.

Niblack W. et al. 1993. The QBIC project: Querying images by content using color, texture, and shape. SPIE 1908, Storage and Retrieval for Image and Video Database:173-187.

Pinkerton, B. 1994. Finding What People Want: Experiences with WebCrawler. The Second International World Wide Web Conference. http:/webcrawler.com/ WebCrawler/WWW94.html.

Salton, G. 1989. *Automatic Text Processing: The transformation, analysis, and retrieval of information by computer.* Addison-Wesley.

Simon, H. 1967. The Sciences of the Artificial, MIT Press.

Simeaton, A., and Quigley, I. 1996. Experiments on Using Semantic Distance Between Words in Image Caption Retrieval. Proc. of the 19th Intl. Conf. on Research and Development in Information Retrieval, Zurich.

Stollnitz, E. J.; DeRose T. D.; and Salesin H. 1995. Wavelets for Computer Graphics: A Primer Part 1.IEEE Computer Graphics and Applications 15(3).

Uriostegui, A. 1996. Digital video indexing applied to one deportive event. (In Spanish) Msc Thesis, CICESE.

Voigt, K. 1995. Reasoning about Change and Uncertainty in Browser Customization. Workshop on AI Applications in Knowledge Navigation and Retrieval, AAAI-95 Fall Symposium Series:136-141.

Wurman, R. S. 1990. Information Anxiety. Bantman Press.

SunMicroSystem. http//www.java.sun.com