

WebSeer: An Image Search Engine for the World Wide Web¹

Charles Frankel, Michael J. Swain, and Vassilis Athitsos

The University of Chicago
Computer Science Department
1100 East 58th Street
Chicago, Illinois 60637

Technical Report 96-14

August 1, 1996

Abstract

Because of the size of the World Wide Web and its inherent lack of structure, finding what one is looking for can be a challenge. *PC-Meter's* March, 1996, survey found that three of the five most visited Web sites were search engines. However, while Web pages typically contain both text *and* images, all the currently available search engines only index text. This paper describes *WebSeer*, a system for locating images on the Web. *WebSeer* uses image content in addition to associated text to index images, presenting the user with a selection that potentially fits her needs.

¹This work was supported in part by ONR contract N00014-93-1-0332 and NSF Grant No. IRI-9210763-A01.

Introduction

The explosive growth of the World Wide Web has proven to be a double-edged sword. While an immense amount of material is now easily accessible on the Web, locating specific information remains a difficult task. An inexperienced user may find it next to impossible to find the information she wants; even an experienced user may miss relevant Web pages. Users searching the World Wide Web have a number of options presently available to them. Lycos, Excite, Alta Vista, and Yahoo! are but a few examples of useful search engines. All of these systems have been designed primarily to find text-based information on the Web. WebSeer is designed to find the *other* major source of information currently available on the Web: images.

In order to find specific images, a method of discovering and indexing their contents must first be developed. The pixels alone are insufficient; some additional information is needed to make it possible to search for specific images. Discovering image content has proven to be an extremely difficult problem. One approach has been to supplement the image content with other types of information associated with the image. Ogle and Stonebraker's Cypress system [1][2] uses information contained in hand-keyed database fields to supplement image content information. Srihari's Piction system [3] uses the captions of newspaper photographs containing human faces to help locate the faces. IBM's QBIC system [4] relies on the user specifying specific visual cues or providing an example image (e.g. a sketch) to begin a query for an image. In Picard and Minka's Foureyes system [5][6] close interaction with a human user supplements information derived from the image content.

WebSeer uses the textual information surrounding an image and the image header to supplement the information derived from analyzing the image content. This additional information is used to create a context in which image analysis algorithms can effectively operate. Image analysis algorithms are then used to classify the image within a taxonomy of types (drawing, photograph, portrait, etc.) and to extract useful semantic information.

Like text search engines, WebSeer does not have to access the original data to respond to a query; all analysis of the image and surrounding text is done off-line during the creation of the database. In this way, WebSeer will be able to give fast query responses to a possibly huge number of users.

The first part of this report describes the types of information that an image search engine such as WebSeer could extract from the Web and how they could be stored in a database. The second part illustrates a number of example queries and how images could be found that fit the queries. The final part details the prototype WebSeer program and the algorithms that have been implemented to date.

How to Find Images on the Web

Information for finding images on the World Wide Web can come from two sources: the relevant text and the image itself. Using information from both sources, a program should be able to successfully retrieve requested images.

Cues from the Text and HTML Source Code

An HTML document is a *structured* document. Understanding the structure of a particular document can reveal valuable information about the images contained on that page. There are several places relevant information about image content may be located within the document. These are ordered by the likelihood of the text being useful in an image search.

1. **Image File Names:** File names contain important clues as to the content of the file. In many cases the full URL of an image (e.g. “<http://www2.int-evry.fr/~hug/images/marylin/mm-trans.gif>”) contains more information than a relative URL (e.g. “[mm-trans.gif](#)”), which may be included in the HTML source. We have found the file name and the name of the top-level directory containing the file are most useful for indexing the content of the image. A caveat is that filenames may be difficult to decipher because of abbreviations used for brevity or out of necessity (e.g. in the FAT file system used in DOS/Windows 3.1).
2. **Image Captions:** Often, images have captions that describe the picture. Although HTML does not have an explicit caption tag, captions are often signaled as text that is located within the same *center* tag as the image (<http://www.lib.uchicago.edu/~mozart/Marilyn/monthly.html>), within the same cell of a table as an image (<http://studentweb.tulane.edu/~llovejo/marylin/>), or within the same paragraph (see Appendix A).
3. **ALT= Text:** Image fields in HTML documents may contain a special ALT= section for alternate text which is displayed in the event that the images can not be loaded. This alternate text may briefly describe the content of the image. For instance, the alternate text for the a photograph of a person may contain simply the name of the person in the photograph. The alternate text may also give hints as to the intended “goal” of the image. For instance, an image containing an ad for Coca-Cola may have the alternate text “Buy Coca-Cola.”
4. **HTML Titles:** The HTML title of the document (which is displayed above the viewed page by the browser, is used for history lists, and so on) often provides information about the content of the images contained within the document. For example, the page <http://tomahawk.cf.ac.uk/~scotw/marylin/page2/page2.html> is entitled “Marilyn Monroe Images - The ventilator scene”, which is highly descriptive of the contents, and more descriptive than any other cues that can be found on the page (see Appendix A).
5. **HyperLinks:** The text of a hyperlink usually gives clues about what the link refers to. In many cases, images are referenced by hyperlinks rather than being imbedded within the text of the page. An example of such a link is “Marilyn JPG #21 (Jumping for joy)”, at

<http://www.netins.net/showcase/agency/marylyn.html>. The same photo is referenced as “Marilyn *jumping for joy*” at http://www.rtmol.stt.it/rtmol/onsite/stars/marylyn/m_imag~1.html, where only the italicized text is the text of the hyperlink. In many cases, text hyperlinks assume the reader understands that the pictures are of Monroe and does not give any information about who is in the picture. These links say “Simply gorgeous”, “Leaning on a banister”, and so on. Consequently, other text in the same sentence or phrase with the hyperlink can also be relevant to deciphering the image content.

6. **Other text:** In addition to those types mentioned above, other text can give cues about the images contained on the page. This text could be used to index the images on the page, with a lower likelihood of reference than the text categories listed above.

Cues from the Image Content

Although it is clear that the context provided by the surrounding text can produce an effective image search engine, examining the images themselves can substantially enrich the search. In this section we will summarize some of the relevant information that can be derived directly from the image content.

The following attributes can be easily obtained from the image header.

- **Grayscale vs. color:** Some may wish to search for only color images. While color images can be converted to grayscale, those in search of black-and-white photography will want the real thing.
- **Image Size:** This is usually useful for those who wish to screen out images smaller than a certain size.
- **File Type (JPEG, GIF, GIF89, etc.):** Some searchers may wish to avoid the compromises in quality introduced by constraints of these formats, especially the limited color palette in the GIF format. GIF89s are animations, and so will often be of separate interest than the other static image formats.
- **File Size:** For those with slow connections or limited RAM or hard drive space, overly large files may be of less interest
- **File Date (when present):** If someone is looking for newer images, the date can be used to screen out images that are certainly older than a certain date.

Obtaining information beyond that listed above generally requires some image analysis. The most useful sorts of image analysis can be grouped into three overlapping categories: 1. classifying the image into one of several types (e.g. photograph, hand-drawing, computer-generated drawing), 2. recognizing image structures (e.g. faces, horizons), and 3. distinguishing specific regions primarily by color/texture attributes (e.g. skin, sky, foliage).

Classifying Images

Creating a taxonomy of image types can facilitate separating images the user is interested in from those that she is not. However, the most important aspect of classification is creating the categories. The authors of an image search engine must either design a system that uses helpful categories, or include a mechanism to allow the user to guide her own search using relevant schema. Ideally, a system would use aspects of both types of categorization.

One important classification which can be made is between images that are photographs and hand-drawn and computer-created drawings. In many cases users may wish to find only photographs of faces, and ignore the many other images that match the search string. Some of these non-photographs may be weeded out because they are too small (e.g. buttons, dividers, arrows). Others, such as image-maps containing text, advertisements, or drawings, must be eliminated on other criteria. Of course, more realistic drawings will be likely to be mistakenly categorized as photographs. Likewise photographs can be thresholded and manipulated in other ways to make them look more like drawings. Nonetheless, we have found that many images can be successfully classified as either drawings or photographs. It may also be possible to separate computer-generated drawings (i.e. those that are artificial-looking, with precisely straight lines and precise color gradients) from hand drawings.

There are many other potentially useful categories. Images can be classified as portraits, which can be further classified by the number of people in the image and the amount of the body shown along with the face (close-up, head-and-shoulders, upper-body, etc.). Images can also be classified as indoor or out, and outdoor images can be classified as landscapes.

Some of these categorizations may be best determined by analyzing global statistics, or the statistics of relatively large portions of the image (e.g. photo vs. drawing). Others require other sorts of analysis, which are described below.

Recognizing Image Structures

There are many types of objects to identify in images, which would aid content-based indexing. We have identified two “objects” that are both important to the typical WebSeer user and within the capabilities of existing computer vision technology. They are:

1. **Faces:** Face-finding algorithms have recently become fairly successful at finding faces in grayscale photographs [7][8]. Faces in photographs are often seen upright, with the subject looking at the camera; the easiest situation for face-finding algorithms. The high quality of many photographs found on the Web, in many cases produced by professionals with careful control of the lighting and high quality film and optics, improves the reliability of face detectors. Combining face detection with color information from skin detection should make a reliable face detector. A primary role of a face detector is to distinguish images of people from other types of images that may show up on the same page. However, face detection can do more. Users may wish to find a portrait of a single person, a group photo, or pictures of a “crowd” containing many faces. Portraits can be close-ups, head-and-shoulders shots, upper-body, or full-body shots; the size and location of the face can differentiate among these choices.

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.