Using Images Posted on the World Wide Web

R. Jeffrey BLAIR

Abstract

The Internet is a collaborative tool that allows people to share computer resources. The World Wide Web in particular is a vast repository of text, images, and other digital files attached to a global copy machine which can be navigated by hyperlinks. This paper discusses some issues concerning the linking and distribution of digital images and the role of person-to-person communication in mediating that process.

Since Vint Cerf and Bob Kahn engineered the expansion of ARPAnet (Advanced Research Projects Agency net) into the Internet in 1973, it has become a popular and powerful educational tool. The World Wide Web in particular offers an almost unlimited amount of material for student researchers (see Blair, 2002). With billions of documents stored on millions of servers throughout the world, the great diversity of people who use and produce webpages, and minimal control of Internet activity, the line between freedom and anarchy is blurred beyond recognition. The various users and producers each have their own individual motives for constructing and accessing webpages. The situation raises some complex issues about the proper use of the posted material, issues that will not be resolved easily or soon. The potential for tension and conflict between the academic, artistic, and commercial sectors of the Internet was forcefully brought home to me a couple years ago when I received demands from an irate webmaster and frantic pleas from university administrators to "remove" photos—actually, direct links to photos—from a student's webpage on my website. While I was happy to comply,

this gave me a personal incentive to take a hard look at the situation from both academic and non-academic perspectives.

Much of the confusion about the proper use of images on the World Wide Web is rooted in its newness. The Advanced Research Projects Agency (ARPA) was created only three months after the launch of Sputnik I to *eliminate inefficient duplication* among the many groups involved in the U.S. government's high-tech research programs (Bryant, 2000). By 1966 attention focused on creating a computer network that would enable people around the country to *share their computer resources*. ARPAnet expanded into a global network—the Internet with file sharing, e-mail, and remote logons. The idea to *store and connect* information in a web-like structure came to Tim Berners-Lee in 1989. He created Hypertext Mark-up Language (HTML) and the system of Uniform Resource Locator (URL) addresses. Four years later Marc Andereessen and Eric Bina of the University of Illinois added two seminal improvements to the system with their browser program called Mosaic: (a) *images* and (b) *hyperlinks* that opened with just a click of the mouse. Over the next two years the Web went from eleventh place in Internet traffic to first, thanks in part to the fact that Jim Clark and Andereessen posted Netscape Navigator on the Web and *gave it away* to non-commercial users. We are only ten years removed from these image/link innovations. It is not surprising that many people still think about the use of images on the World Wide Web as if they were images in a book. The Web, however, is not just a super-sized library. It is also a copy machine.

Global Copy Machine

The World Wide Web is a *global copy machine*. This is no accident. It was purposely designed that way to allow and encourage *copying and sharing* of digital resources. Instead of buttons this copy machine utilizes *URL addresses* and the Information Super Highway to distribute digital copies around the world. People at computers "push the copy buttons" when they type URL addresses

<div align="center">

http://www.server/...folders.../document

</div>

into the address box of their web browsers.

Webmasters post original *source documents*, including html documents and image documents, on special computers known as servers. These source documents

<div align="center">(210)</div>

are the blueprints for webpages. The servers that store these original source documents can transmit copies to any computer connected to the Internet. The receiving computer is called a client. Browser software on the client computer reads an html document and any related (jpg and gif) image documents and *combines the text and images* into a single browser image called a *webpage*. All the text is contained in the html document and each separate image is contained in its own distinct source document. Suppose you view on your computer a webpage that has text and three images. Then you have downloaded copies of four source documents: one for the text and three separate source documents for the three images. It seems like you are viewing a single image out somewhere in cyberspace. You may imagine that you are viewing an image from another computer, but in reality every webpage you see "on the Web" has been *created on your computer from copied source documents.*

Mixing and Matching Images

An html document contains not only the webpage text, but also the URL locations of all links and all images displayed on the webpage. Any computer that opens the html document will *automatically request* copies of the image documents at the specified URL locations. That is why connections that produce embedded images are sometimes referred to as *direct links*. The standard hyperlink, that seminal feature that makes the World Wide Web a web, initiates a request for copies of the linked document *if and only if* the user chooses to click on the link. Unlike these *click links*, direct links do not involve any specific action on the part of the viewer. This is the way all images are loaded onto a webpage even when they come from the same website as the html document. Each image document arrives after the html document, and the image and html text are then combined *on the client computer.* Unlike printed media, *all webpage images are copies* that have been "pasted" into their webpages. Although images appear to be an integral part of the webpage, they are posted on the Web as *separate and distinct entities* with their own URL addresses.

<div align="center">

http://www.server/...folder.../document.jpg

OR

http://www.server/...folder.../document.gif

</div>

When you view a webpage, only a single URL address appears in your

browser's address box—the address of the html document. This contributes to the mistaken feeling that the webpage is a single entity. Image documents, however, do not need to be combined with an html document to be seen. They can be viewed individually simply by typing their URL address into the browser address box. The image will then appear all by itself in the browser window in its original size.

The fact that *any* image can be displayed on *any* webpage or linked to it creates a great deal of flexibility for the entire Internet community. The owner of an image—the person on whose website the document is posted—can insert that image into *any number* of his own webpages by writing that URL address into the corresponding html documents. The same holds true for other webpage authors. Anyone can write an html document which will download and display the image. This, as stated above, is exactly what the Internet was intentionally designed to do—share computer resources. Hyperlinks have made it simple and convenient to make these connections.

They have also made it simple and easy to break connections. That is why there are so many *dead links* on the World Wide Web (Wales et al., 2005). A webmaster links to an external webpage or image without the need for any coordination or communication. If, however, as often happens the external webpage or image is moved, removed, or renamed that hyperlink fails to work. It is here that directly linked images have a great advantage over click links. When direct links go dead the image disappears from the webpage leaving behind an empty box. Because it is visible anyone who views the webpage knows immediately what has happened, and corrective action can be taken in a timely manner.

These are the essential characteristics of the World Wide Web: it is both a *decentralized* digital *library* of html and image documents and a global *copy machine*. Documents can be *linked* and images *inserted* with great ease. Ideally the process is a symbiotic one in which document owners and webmasters interact in a cooperative, productive manner to build a continually expanding and improved Web of resources available throughout the world. The world is not an ideal one, however, and issues come up. I would now like to explore four of them: control, credit, transmission costs, and communication.

Access and Control

The Internet in general and the World Wide Web in particular are maintained and used by such a great diversity of people that it may be the "most ... extensive medium of personal exchange" in history (Wales et al., 2005). There is an overtly commercial sector, which seeks a direct profit for providing access to its documents. The economic law of supply and demand dictates that such access be restricted. The same is true of websites that handle information of a private nature or where users may be engaged in fierce competition with each other, such as college and financial aid applications. This paper will not concern itself with such websites, except to note that they employ a number of successful strategies to *limit access* and exert control over their materials. Such websites typically use a system of usernames and passwords. Other strategies include the blocking of direct links from external html documents and disabling the print function on the client browser.

Although webmasters and the managers of their servers can limit access to the documents on their website, most *choose not* to do so. They *want as many people* as possible around the world to view the text and images they provide. That is precisely why they posted them on the *World Wide* Web. Some web authors go so far as to include counters in their html documents in order to measure the traffic and then boast of how many hits their webpage gets, how popular it is. The problem for those who wish to retain control of their materials is that to the extent that you provide access, you give up control. There is a trade-off between the two. Artists, including photographers and writers, are particularly sensitive and vulnerable. They understandably want to retain artistic control of their creations.

Comparison with a typical printed product, such as a book, will illustrate what a great difference the Internet makes. Books are printed in limited numbers. People pay for them, read them, and sometimes pass them on to friends. Some books are put in libraries, where many people can borrow a single book. Most libraries have copy machines, and it's not unusual for people to copy sections of some books. Copying an entire book, however, is time consuming and costly as well as obnoxious. It is almost always easier and less expensive to buy the book.

The non-commercial webpages under discussion in this article are posted on a global copy machine for any and all to access free of charge other than fees paid to an Internet service provider. Anyone who has viewed a webpage has requested

and received a temporary copy of its html document and of each image it displays. Making more permanent copies of the images or text is often as easy as click-and-drag or copy-and-paste. There is virtually no economic incentive or barriers involved since the text and images are freely available on the Internet. *Any computer anywhere in the world* that is connected to the Internet can copy such source documents. Unlike a library of printed material, you do not actually borrow or directly view a webpage. Instead, you get a copy of its blueprint—the source documents; the webpage is constructed on your computer; and you never return the copies. It is as if a library printed up copies of its own books and gave them away for free. Copies of documents on Internet servers are transmitted directly from the websites of owners to client computers around the world. Let us call this the *primary mode of distribution.*

The primary mode occurs as a *chain of events* that ends with the *original website*—the document owner's website—transmitting a copy of the original source document to a client. We will call him the *primary client*, because he receives the document from the original website. This chain of events, furthermore, always begins with the primary client. There are three possible routes the chain may take according to whether the request for the document is initiated with (a) a typed URL, (b) a click link, or (c) a direct link. When a primary client types out the URL address, he probably has a good idea of what document or kind of document he is requesting. The search for webpages often begins by going to a search engine, like Yahoo or Google, with a URL address that is easy to remember,

<div align="center">

www.google.com

OR

www.yahoo co.jp.

</div>

After that it usually continues with a series of click links each of which leads to (ie. downloads) a *single* html document. The client makes a conscious choice to click words, phrases, or images contained in an html document. The words, phrases, or images will often give them a hint as to the document or kind of document that they are requesting. A click link to an html document may also download *multiple* directly linked image documents. This is an automatic consequence of viewing that html document. Where a click or direct link leads, that is, which documents it tries to download onto the client computer, of course, is decided not by the primary client or by the original document owner, but by a third party—the author of the html

document. Yet document owners are still in *full control* of this mode of distribution. The website can distinguish between a typed request and a link request. Websites can and sometime do refuse external link requests for documents. The owner can impose limits on this access or even remove some documents and thereby end access to them completely. What owners cannot control is the *secondary mode of distribution*.

The secondary mode occurs as a chain of events that does not include the original website. A primary client can transmit document copies to other client computers via e-mail attachments. These secondary clients, or any primary client, can also upload these documents onto websites other than the original, where they will be available to the entire cyberworld. As long as access through the primary mode is *unlimited*, access through a secondary mode merely serves to augment the distribution of those documents to people who already have access. Rather than *unauthorized access* the main issue is acknowledgment.

Credit to Owners

The owners of original image documents, their websites, and the photographers or artists that created those images deserve to be acknowledged. Traditional print media usually accomplishes this with a special section in publications that use borrowed images. That credit section lists the original published sources and creators of the images. Readers who are interested in tracking down the source material can search in libraries and contact publishers. This cumbersome process has been streamlined on the Internet with the use of click links. Images can be reverse linked—linked from a webpage back to the original image document—or, alternatively, to an html document on the original website that includes the image or acknowledgments and details about it. The URL address that appears in the browser address box will identify the original website. For the sake of readers of a printed webpage it is a good idea to also include the URL addresses of all images in an image credit section.

The original webmaster can facilitate or enhance the availability of proper credit information by gathering images into a single folder or series of folders, each with an html index document that provides appropriate details concerning the folder's images. Then when a viewer truncates back to that folder, the desired information will automatically appear. The use of links, folders, and html index

(215)

documents in this way allows the World Wide Web to centralize the acknowledgment process. It eliminates the need for multiple presentations of similar information and gives the original website control over both content and form of presentation.

Even if a webpage fails to link an image to its original or provide any information in its text about image sources, viewers can still track it down in the html source document as long as a copy of the image document has not been uploaded to a different website. The direct link which calls up each image specifies the URL address. Although it may be buried away fairly deep in the html code, with sharp eyes and a little practice almost anyone can learn to recover the URL of the original image document. In this sense, direct links always provide some form of acknowdgement to the website that posts an image. Webmasters who wish to make sure that viewers are aware of their contribution to the World Wide Web even when image documents have been uploaded to different websites, on the other hand, may need to print their URL address on their images.

Burden of Transmission

While directly linked images provide at least a minimal acknowledgment of the linked website, they also place a burden on the that site's bandwidth. Some people complain that this is a form of stealing in its own right. On the contrary, this clearly shows that the original website is distributing copies of its own image document (primary mode). Traffic to and from a website does, indeed, increase each time a document is requested and transmitted. This is an inevitable consequence of any communication between computers. In theory, however, the communication is of *benefit to all parties*—the party that wants to view the image, the one that wants to display it, and the one that posts it on the Web for display.

These transmission costs are like the service charge for a cashier's check. Some banks ask their customers to pay a "service fee" when they issue them a cashier's check, usually for a large amount of money. Yet when the customer chooses a cash withdrawal rather than pay, these same banks inevitably waive the fee, because this "service" is in their interest as well as the customer's. Similarly, although websites provide a service to whoever views their material, this service must also have been in the interest of the webmaster and any contributors in order for them to have provided unlimited and free access to it in the first place. Thus the burden of

increased computer transmissions is offset or balanced by the additional promotion those images receive.

The structure of the World Wide *Web* is based upon links. That is the mechanism by which people navigate from a search engine to the webpages they want to view. It is the links on search engines and other html documents that advertise and promote webpages and thereby make their presence felt. A document free floating in cyberspace, without any links to it, would be invisible, virtually non-existent. Thus links are necessary. In addition, their potential for expanding traffic to any specific document is very powerful. Each link funnels a percentage of a webpage's traffic to the next webpage downstream. For direct links the percentage is 100%. For click links it is the percentage of viewers that choose to click on a particular link. Click links can send a stream of computer traffic through a series of webpages converging each time with other similar streams to eventually form a raging river of traffic to some well-connected image document. It is this *geometric progression of links* radiating upstream from a document that gives the Web its power, *empowers* popular webpages and images, and can sometimes *overpowers* them, even though the burden for each single request and transmission is not great and not affected either by the route it takes or by the nature of the final connection—typed, click link, or a direct link.

While the owners of documents are necessarily saddled with the burden of primary mode transmission, any distribution in a secondary mode will help them avoid those transmission costs. In other words, there is a trade off between control of access and transmission costs. Owners of popular image and html documents may be willing, even eager, to sacrifice control over their distribution in order to minimize their transmission costs. For this very reason whole websites are sometimes copied and loaded on other servers and the result called a *mirror website*. This can be accomplished, on a smaller scale, by loading the copy of a single image document onto a new website. The copy can then be embedded in one or more webpages using direct links to the new website, while connecting the embedded images to the original website with click links, and thus providing acknowledgment. The percentage of traffic flowing through the final connection will be reduced from 100% to the percentage of people that choose to click on the reverse link.

Loading someone else's documents on your own website, of course, deprives them of control over their distribution and thus requires their permission. While

(217)

this should reduce the ongoing operational burden of computer transmission, it imposes a communicative burden in two directions—one upon the person requesting such permission and another upon the person receiving, considering, and replying to those requests.

Burden of Communication

We have established that the World Wide Web is a global copy machine for digital information and discussed some of the possible ways these copies might be shared. Realizing that securing permission from image owners is desirable even when it may not be mandatory, we now turn our attention to how that might best be accomplished—communication. It sounds so easy. Image owners could simply post the terms of use on their website along with an e-mail address for image link permission requests. Whoever makes direct links to images at another site could alert the original website to the existence of the connecting webpage, so that they could inspect it, evaluate the manner of linkage, and respond. In theory, this communication process should be accomplished within a few days, long before the rest of the cyberworld has time to find and make their own connections to the new webpage. Unfortunately the reality of the Information Super Highway is much more complicated.

The Information Super Highway, after all, not only gives us instant access to information, it also bombards us with it, spam and viruses included. The floodgates are open; the in-box fills up; and we have to deal with it. Opportunity may be boundless, but the competition is fierce. Since we *can* do more, we *must* do more. Our lives become frenzied trying to keep up. Websites are continually under construction. Links go dead for want of maintenance. It is not surprising, therefore, that many websites lack any contact information or that some posted e-mail addresses are out of date. Sometimes this failure to provide an e-mail contact is a purposeful strategy to protect busy webmasters from what could be a deluge of e-mail, legitimate or otherwise. If they do not want to discourage all contact, a webmaster may provide a postal address or a special form submission webpage. This filters contact with the cyberworld down to those who are willing to make the extra effort.

Other factors may also discourage people who would like to make contact. On websites with numerous webpages, contact information may not be on each

and every page. Although, in such cases, links are often provided to a page with the information, more and more powerful computers manage to squeeze more and more information on to individual webpages. This information overload can make it difficult to find the appropriate link and/or to find the right contact information even when viewing the correct webpage. This is especially true if the person is not a native speaker of the language used on the image owner's site.

The search for contact information is one obstacle to communication, but it can usually be overcome. It is mostly a matter of having enough time and making the effort. The next obstacle is more difficult to surmount. After link permission has been requested, the most common response in my experience is silence—no response. We can only speculate on the possible causes: (a) lack of concern about linkage, (b) lack of time to evaluate requests and respond, (c) language barriers. Most people who have responded to my link requests have responded positively, granting permission. Aware that the structure of the World Wide Web in based on links, many people probably do not perceive a need to grant permission for them. Alternatively they may just be too busy to give serious attention to the request. Any language barrier will compound the problems of understanding the request, evaluating the webpage that wants to link, and formulating a response. I used to send English requests to Japanese websites whose images my Japanese students had chosen to illustrate their research papers until an irate Japanese company complained to my university about my sending requests in *English*. It took me a couple months to work up a bilingual version of the request. I have posted it on the Web ‹www.aichi-gakuin.ac.jp/~jeffreyb/permission.html› and now refer all Japanese webmasters to that webpage if they prefer Japanese.

The point to note here is that communication in the real world, even with the Internet, represents a real burden to both parties, the website requesting permission and the website that has to evaluate the request and respond. Some webmasters try to minimize their burden by posting pseudo-legal notices and/or giving formulaic responses to requests. Though this may be a viable tactic to try to preserve whatever rights exist for materials knowingly placed in a copy machine universally accessible to the cyber *public*, the actual legal implications of these notices are extremely difficult to assess. As a global network the World Wide Web is potentially subject to a multitude of uncoordinated and possibly conflicting national laws. How have those laws have been interpreted, reconciled, and applied by national or international courts is unclear. While I would be happy to hear from anyone who is privy to

(219)

such information, that line of inquiry will have to remain beyond the scope of this paper.

Conclusions

The Internet was created as a collaborative tool to allow people to share computer resources. Within this general framework the World Wide Web was specifically designed as a global copy machine attached to an extensive decentralized library of digital files. There are three roles that an individual can assume in the sharing of documents: (a) the owner/manager of posted documents, (b) the viewer, or (c) an author of a linking document. The owner posts original documents—html or image documents, for instance—on a server. It is his own website which *makes and transmits copies* of these document upon the request of client computers. After a client computer receives copies of an html document and its accompanying image documents, it *combines them* into a single webpage and displays it for the viewer to see. Viewers usually find documents by clicking through a series of hyperlinks from one html document to another. Copies of each such document and its image documents are *stored on the client computer.* The author of a linking html document, neither copies nor stores any documents, but simply facilitates the process of locating and retrieving them by establishing routes along which viewers navigate to gain access.

Documents can be shared in one of four ways: (1) a viewer requests a copy and views a document by typing or pasting its URL address into his browser, with no involvement by a third party, (2) viewer access through a click link, (3) viewer access to and display of an image via a direct link, and (4) the posting or sending of copied documents. The first poses no problems, since our discussion is limited to websites which authorize access to everyone. The second will tend to increase the traffic to linked documents, but only to a *single* document with the *deliberate consent* of the viewer as he clicks the hyperlink. The third may increase traffic to *multiple* image documents with only the *implied consent* of the viewer, who has not clicked any specific links, only chosen to view the linking document. It may also take away the owner's artistic control over the way an image is displayed by integrating it into another author's webpage. All three, however, represent primary modes of distribution, in which the owner's own website makes and transmits all document copies. The owner remains in control of any further distribution. The fourth method does

take away the owner's control of document distribution, but in exchange, relieves him of the burden of transmission.

The sharing of computer resources should be of *mutual benefit* to all three parties involved in the process—the owner of a linked document, the author of a linking document, and the viewer (who views both documents, often as a single webpage). *Communication* between the linking author and the linked owner is essential to this end. It should flow in *both directions* with the burden of responsibility moving back and forth between them.

Acknowledgments

Points of Contact

Any comments on this article will be welcomed and should be mailed to the author at Aichi Gakuin University, Junior College Division, 1-100 Kusumoto-cho, Chikusa-ku, Nagoya, Japan 456-0037 or e-mailed to jeffreyb@dpc.aichi-gakuin.ac.jp. Other papers and works in progress may be accessed at http://www.aichi-gakuin.ac.jp/~jeffreyb/research/index.html.

References

Blair, J. (2002). Research and process writing on the Internet. *Faculty Journal of the Jr. College Division of Aichi Gakuin University*, 10, pp. 94–106.

Bryant, S. (2000). *The Story of the Internet*. Edinburgh: Pearson Education.

Wales, J. et al. (Eds., 2005). World Wide Web. *Wikipedia*. Posted at http://en.wikipedia.org/wiki/World_wide_web.