Deep linking provides a means of optimizing the information extracted from existing third-party websites in general, and from search applications in particular. Various means of deep linking are introduced such as URL editing, using search forms, and placing a man in the middle. These are the very techniques that are used by many of the tools on the One-Step website (http://stevemorse.org).

Then the tables are turned and methods of blocking others from deep linking to your website are described. The legalities of deep linking are also discussed.

The purpose is not to make you an expert at improving other people's websites. Rather it is to expose you to the techniques that were used on the One-Step website and give you a better appreciation for what is there and how to use it. As such, no knowledge of webpage programming is required. However, if you have such knowledge you will be able to apply the ideas presented here yourself.

URL EDITING

If you go to google and do a search for genealogy, you will be taken to a results page listing all of the 35 million sites involved with genealogy. And the web address (known as the URL) of that google results page is:

http://www.google.com/search?hl=en&q=genealogy

We note that the URL contains a question mark followed by items like hl=en and q=genealogy. We can take an educated guess and conclude that q=genealogy is specifying that we are searching for the word “genealogy.” But what does the hl=en mean?

One way to find out is to modify the web address so it contains something other than en. In particular, we can change the en to fr. When we do so, we notice that the google results page that was all in English is now all in French. So hl is specifying the language – it is telling google what language we want the results to be displayed in.

What we have done is edit some of the parameters passed to the google search engine. The fact that parameter q refers to the item being sought and the parameter hl refers to the language is something unique to google. Each website will have its own parameters, and each parameter will have its own allowable set of values. But if we can figure out what the parameters and associated values are for a particular website, we can manually edit the URL line to make the site do some unexpected things, some of which might be very useful.
This URL editing is something that anyone can do, and you can have fun playing around with. The other techniques presented in this paper are not intended for you to try at home unless you have a detailed knowledge of Internet protocols.

USING A SEARCH FORM

URL editing can be tedious and is error prone. It is certainly not the way one website deep-links to another. Instead a deep-linking website will present a search form that has the capability of generating the URL with all the desired parameters.

As an example, the site might present a form that has two fields. One field is a string to be searched for and the other is the language that the results are to be displayed in. And the form has a submit button. When the submit button is pressed, your browser will automatically generate the appropriate URL with the hl and q parameters having the values you entered on the form.

SEARCH FORM WITH HIDDEN PARAMETERS

Although all search forms generate parameters from the values you enter on the form, not all of them make the parameters visible in the URL. Of course those parameters need to be sent to the website that you are linking to. But they need not be sent along with the URL. Instead they are sent down the Internet wires along with the URL, but in such a way that they are not made visible to you. In this case you can still create a form to do such deep linking. But you cannot do it with URL editing.

MAN IN THE MIDDLE

What we’ve seen up until now has been rather static. Whatever results the website sends back are displayed on the screen as is. By contrast, a man-in-the-middle provides more dynamic capabilities. It can modify the display in any way it sees fit. For example, it can change the search results so that every person named Jack is changed to John. It can filter the results so that nobody named Fred is shown. It can reformat the display of the results. And it can display more results per page.

Note that the man in the middle is not really a man. It is a website that gets in between the person making the request and the final website that receives the request.

How exactly does the man in the middle approach work. Let’s say you are sitting at your compute in California. And you are filling out a search form that will be submitted to a genealogy website in Utah. Normally your search request goes from California to Utah, the website in Utah does the search and generates a results page, and that results page is transmitted from Utah back to California. Now suppose we create a man-in-the-middle and locate it in New York. We change the search form so that instead of submitting to the website in Utah, it submits to the man in the middle in New York. The man in the middle will take the same request that it received, and redirect it to the website in Utah. The website in Utah will perform the search and return the results to the requester, which
in this case is the man in the middle in New York. The man will get the results, modify it in whatever way it was programmed to do so, and then send the modified results back to the person in California who made the original request.

The One-Step website has many examples that use the man in the middle approach. For example, there are One-Step tools lets you look up phonenumbers in the Israeli telephone directory. It does so by interacting with a website in Israel. That website is entirely in Hebrew. So the man in the middle converts your English request to a Hebrew one, sends the Hebrew request to the website in Israel, gets back the answers in Hebrew, and converts the Hebrew results to English before displaying it on your screen.

**COMPLETE TAKEOVER**

The man in the middle is a one-time thing. The search form takes the user to the man, who in turn does his thing, and eventually returns his results to the user. After that the man is no longer in the picture.

A complete takeover involves putting a man in the middle and keeping him there. As before, the search form takes the user to the man, and the man fetches the first page from the actual website. In addition to doing the modifications described above, the man can also change all links on the page so they point back to the man rather than to the actual website. By doing so, the man will stay in the middle for all future accesses to that website.

**HOW TO BLOCK DEEP LINKING**

Let’s switch hats now and suppose we have a website that we don’t want others to deep link to. After all, we worked hard to obtain our data and we don’t want other websites gaining access to it. Remember the golden rule – Do not let others do unto you as you would do unto them!

There are several things that a website can do to prevent others from deep linking to it. One way is to use the so-called “referrer field” so determine how the request was referred to our site and to block if the referrer was not a site we are familiar with. Another is to check for specific cookies since a site that we do not know cannot set a cookie for us to read. And a third is to request the user to read the characters in an image before we send him back any results.

**DEFEATING THE DEEP-LINKING BLOCK**

Now let’s put our original hat on again, and continue deep linking to other websites. We’ve come across a website that has installed a technique to block us, as described in the previous section. What do we do?
Is it possible to defeat the deep-linking block? The answer is yes. So how do you do it? Unfortunately if I told you I would have to kill you. I don’t want anyone defeating the deep-linking blocking that I’ve installed on my One-Step website.

IS DEEP LINKING LEGAL

With all the fuss over deep linking and websites trying to block it, the obvious questions comes about as to whether deep linking is legal. The entire World Wide Web is based on the concept of deep linking, and without it the web would become a very sterile place.

I’m not a lawyer, so what I present here should not be construed as legal opinion. But it is the result of some of my own research on the subject.

The first example of a deep-linking case that I could find occurred in Scotland in 1997. It involved two newspaper websites – the Shetland Times and the Shetland News. The Times had actual news stories. The News had links to the stories on the Times website. The case went to court. The court banned the links based on the fact that the News was using the Times headline verbatim. This was a copyright violation. By giving the verdict in this manner, the court did not have to rule on the deep-linking issue.

The next case I found involved Ticketmaster and Microsoft. Microsoft operated a website called Sidewalk that was a recreational guide to various cities. It offered the user links to Ticketmaster to buy tickets for events in that city. The links drove traffic to Ticketmaster, giving them more sales. But Ticketmaster ignored that fact, and decided to sue. They were more interested in having their day in court than in making money. But Microsoft has much deeper pockets than Ticketmaster, so the case was eventually settled out of court and the terms of the settlement were sealed. No judicial ruling was issued.

Ticketmaster wasn’t happy about this, so they decided to try again. This time the focuses on Tickets.com, who was doing exactly the same thing that Microsoft did. And Tickets.com’s pockets weren’t as deep as Microsoft’s. So now the case went to court, and the court ruled against Ticketmaster. The court’s ruling included the following statements:

“Hyperlinking does not itself involve a violation of the copyright act since no copying is involved.”

“There is no deception in what is happening.”

As I said before, I’m not a lawyer. But the conclusion from this court statement seems obvious to me.