# Mini-Googles: The Benefits of Search Roaming

*Joe McCormack*
*University of Advancing Technology*

By adding related websites to your search engine, you inherently increase the usefulness of your website for your audience.

The ability to search other websites from your own website is an attractive feature for web developers and content owners. One benefit is that you can expose related websites and data to your audience. This can help establish a website as a search launch point for its users by enabling them to search related sites. This could also increase the likelihood that your users will return to your website in the future when they are seeking content with which your website is aligned—even if you don't necessarily have what your audience is looking for, one of the websites that you allow your search engine to search may. As marketing has proven, the more times an individual is exposed to an interesting and/or useful website, the more likely they are to return to it in the future and spend more time on it. In a sales model, this improves the likelihood of generating a sale because you are able to repeatedly expose what you are selling to the same targeted individual.

This method is called *search roaming*. It is a cross-website search method that allows you to use your website's search engine to search other websites of your choosing in real time. Search roaming opens up a myriad of possible uses and benefits. Think of search roaming as being "mini-Googles" across the Internet.

### Regulating What Is Searched

A website is geared to serve a specific type of audience. For obvious reasons, a website that has content about rocket engines is not going to have content about learning how to quilt; because the site is about rocket engines it will serve the type of audience interested in rocket engines, rocket technology researchers, developers and engineers, etc. With that niche, you could introduce related content to your audience (in many cases you will not have everything that everyone else has in a given niche or market). By adding related websites to your search engine, you

inherently increase the usefulness of your website for your audience. Additionally, the websites that you include in your search roaming community benefit from the targeted traffic that you generate for them. In turn, your website can become associated as an "authority" on the subject and a start point for that type of audience. This gives you more opportunities to expose your website to your audience and, the more times an individual is exposed to your website, the more likely they are to spend more time on your site, explore it and trust its content.

### Search Roaming Communities

How can you use cross-site searching to establish a searching community or "ring" (remember the web-ring craze of the late 1990s?)? Provided that each website has search roaming capacity, each website could set up their search engine to search other websites. Collectively, since everyone is connected to everyone else's search engine, the websites in that community share the audience that all of the websites generate. All of the websites gain more exposure which, more than likely, a single website on its own would not be able to generate.

### Regulating What Is Searched On Your Site

In order to control what is being searched on your own website, an index file (similar to a robots.txt file) should be created. This text file specifies which web pages on your website can be searched by others. The layout is very simple, with one path or filename per line encapsulated with **<p></p>** (which signifies "page" not paragraph, as in typical HTML code). For the purposes of unifying what the name of the index file would be, it is typically named **cwpsrsp.txt** and placed in the root of your website. Here is an example of what the cwpsrsp.txt file could contain:

```
<p>index.html</p>
<p>folder/subfolder/somefile.asp</p>
<p>folder/file.php</p>
<p>cgi-bin/page.cgi</p>
<p>simple.txt</p>
<p>data.aspx</p>
<p>a_page.cfm</p>
```

### Search Roaming In Real Time

Searching other websites in real time and having an index to follow makes the search time faster and guarantees that your search results will be instantly updated without any extra effort on your part when content changes on other websites. However, the ability to use real-time searching is limited by several factors: system overhead of your web server and the web servers that are hosting the target websites, the size of the pipe through which data flows, and the path your search engine is taking through the Internet to index data from target websites (so that the data can be searched). If you are searching websites within your own network, for example, response time of the real-time indexing and search process will be higher than if you had a website in California that was indexing and searching websites in Australia, Hawaii and New York simultaneously.

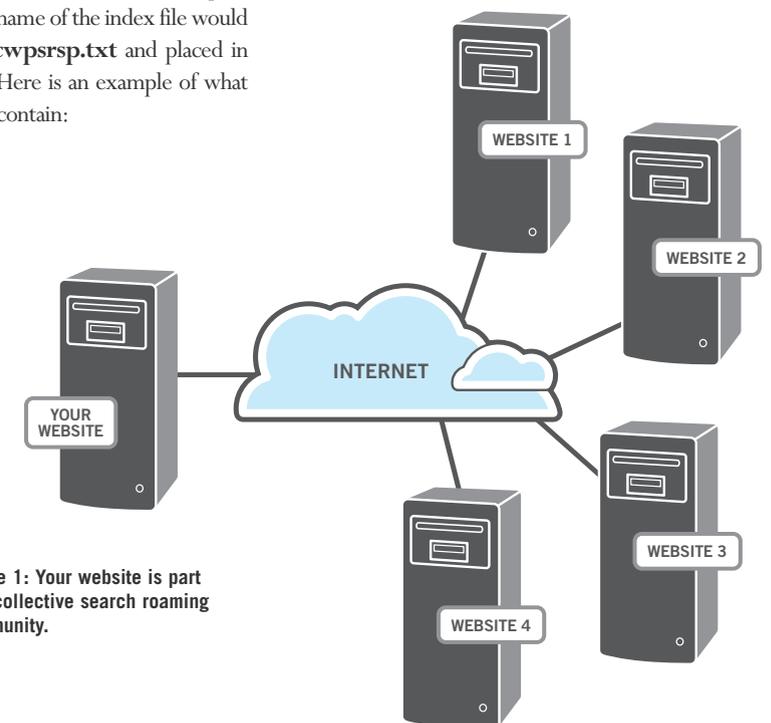Figures 1-5 demonstrate a search roaming scenario in real time.



**Figure 1: Your website is part of a collective search roaming community.**

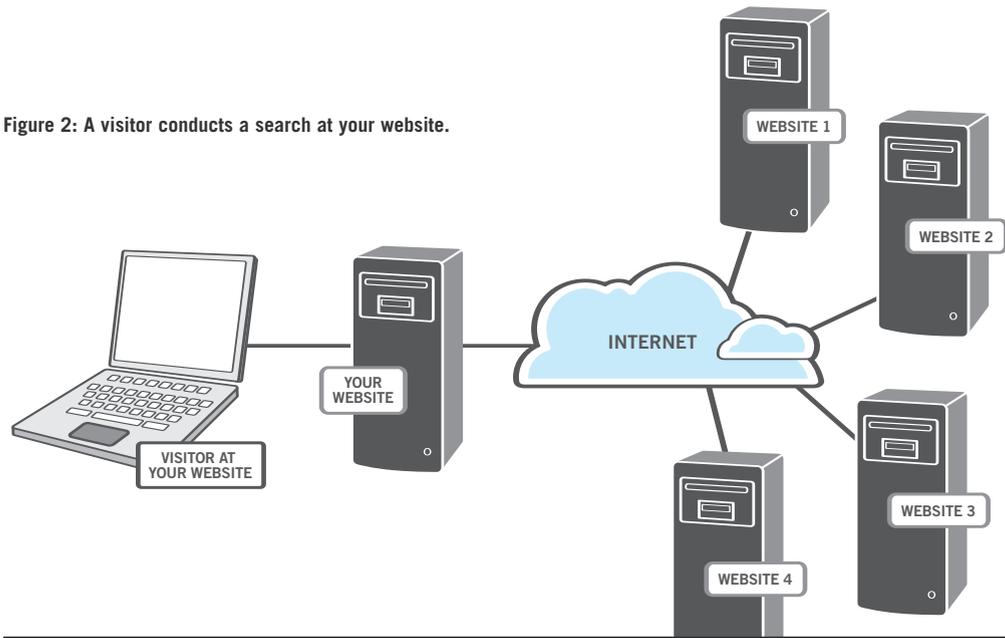Figure 2: A visitor conducts a search at your website.


Figure 3: Your website (hosting the query) asks if any of the other websites in your search roaming community have what the visitor is searching for.
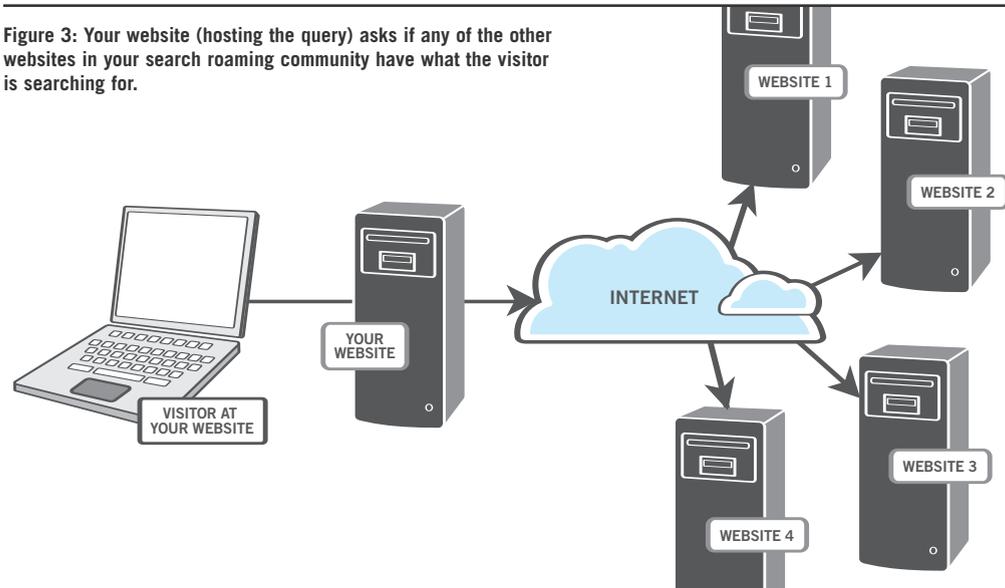

Figure 4: Those websites which have what your visitor is searching for respond to your website.
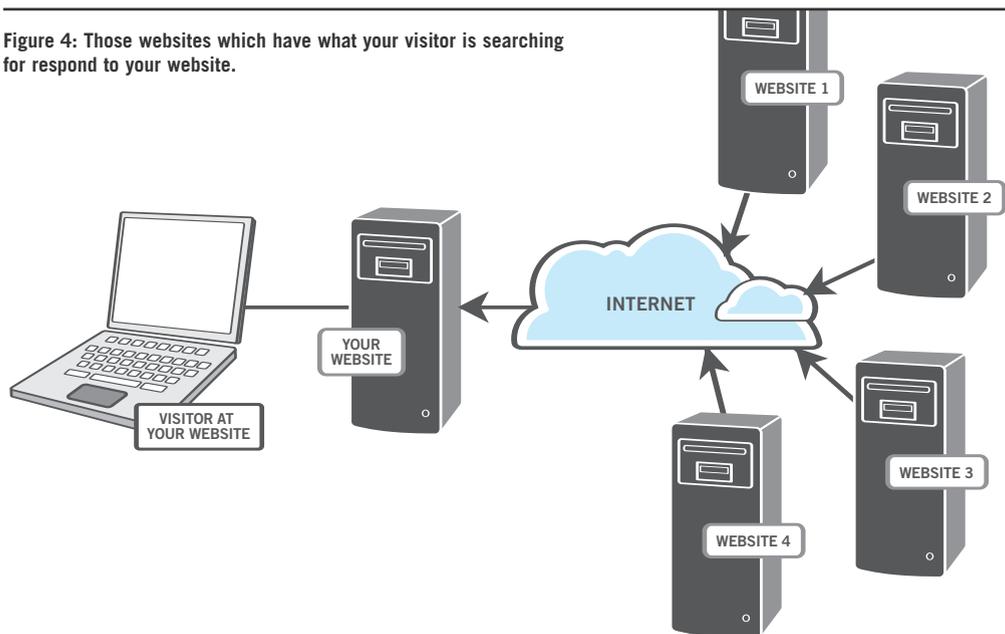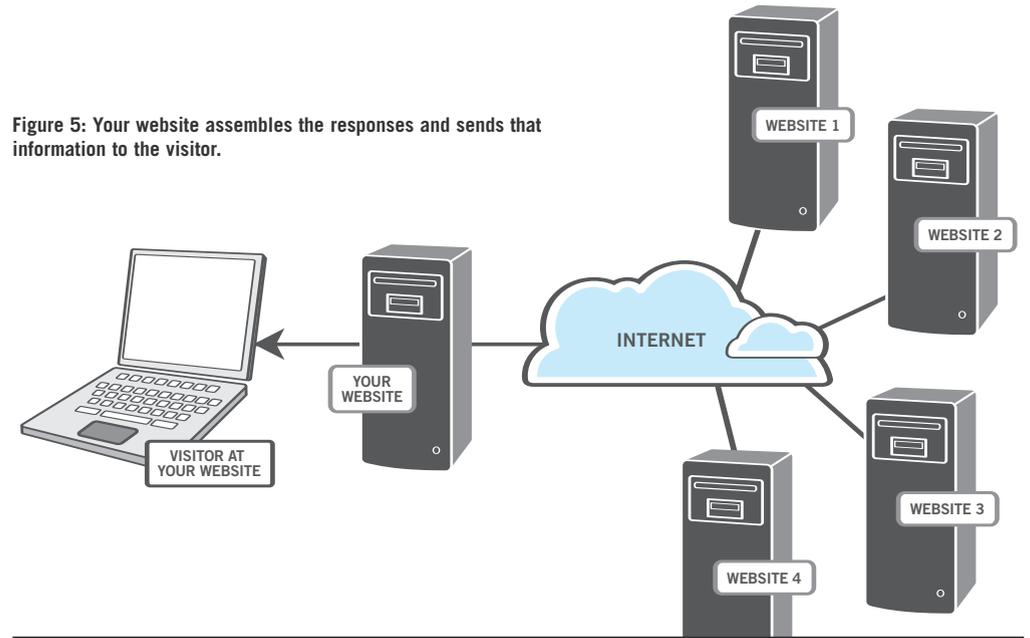

Figure 5: Your website assembles the responses and sends that information to the visitor.

The work-around for not being able to perform searches in real time is to have an index on your website that can be updated every few hours or once per day by an index bot that runs on an automated schedule (such as a cron job on an Apache web server). The index on your website would contain the links to the other sites as well as the content you wish to save from the web pages. Then, your search engine could merely search the local index, thereby reducing the response time of your search engine's search results.

## How Search Roaming Works Using Indexed Data

The process of using indexed data from multiple websites within the search roaming community is fairly simple: add a website's domain name into the search engine indexer.

The search engine indexer will contact the website and look for **cwpsrsp.txt** (which is contained in the root folder of the website) in order to determine which web pages that it is *legally* allowed to search. Each page in the index file will be recorded by the search engine indexer (this may vary based on what you want to record, whether it is metadata, page content or word patterns or occurrences).

## Search Roaming In "Cached Time"

Using a search index, or cache, on your website will greatly enhance the speed at which searching multiple websites is conducted because additional overhead such as bandwidth and server response times of websites in your community are not issues. In figures 6-8 below, you can see how using an index can get around these issues.
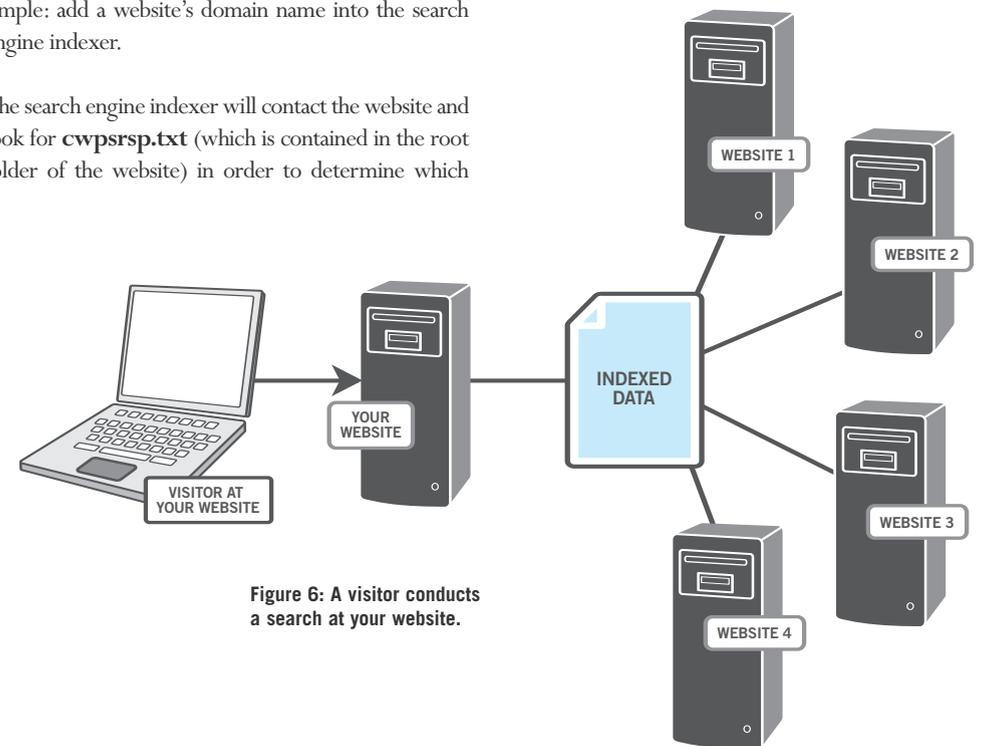
Figure 6: A visitor conducts a search at your website.

**Figure 7: Your website searches the indexed data which contains the authorized data from each website in your community.**
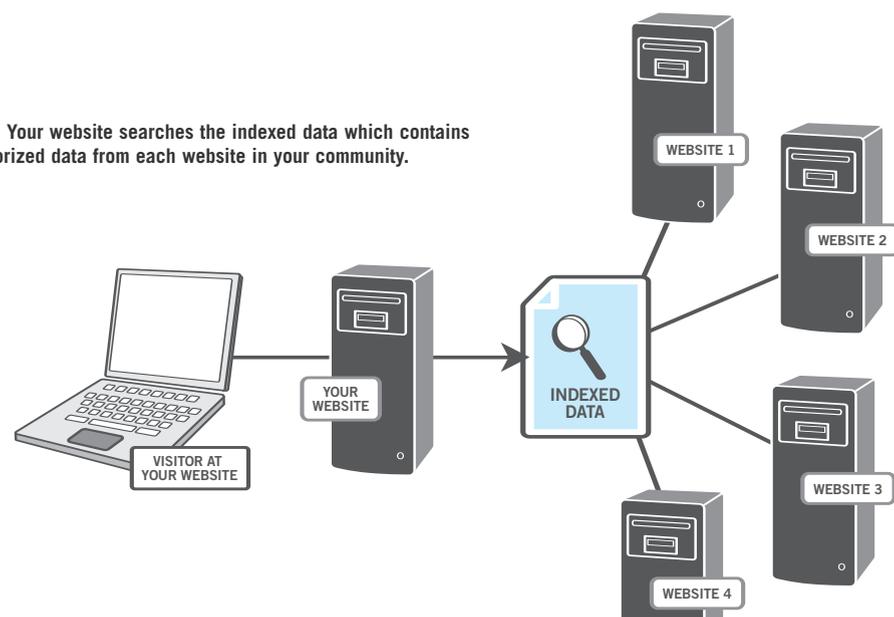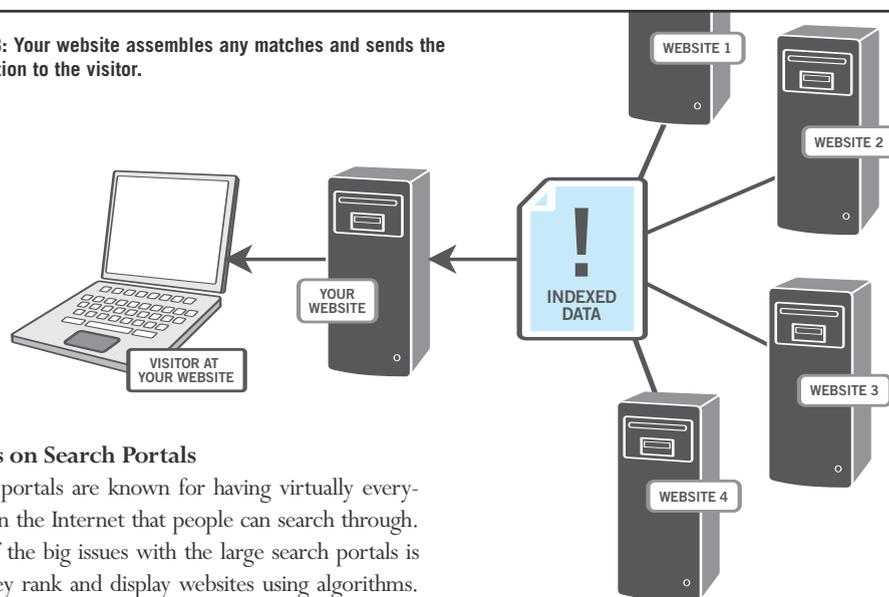


**Figure 8: Your website assembles any matches and sends the information to the visitor.**

# The ESRB Rating System: A Gamer's Perspective

*Jonathan S. Harbour*
*University of Advancing Technology*

**Research and anecdotal evidence show that the potential for harm from video games is much greater than previously understood.**

Video game publishers are releasing games today that are filled with intense scenes of violence, sex and profanity, games that are being played by young children and teenagers in ever-growing numbers. Why is this a problem? Because video games are beginning to look more real due to advanced 3D graphics technology. Thus, playing a realistic game is similar to a real-world situation. That the US Army is using first-person shooters for recruitment and training is a telling point.

The National Institute on Media and the Family succinctly describes the problem with the Entertainment Software Rating Board (ESRB) ratings system in their MediaWise Video Game Report Card:

> While the industry is making the same efforts to protect children it has over the past few years, research and anecdotal evidence show that the potential for harm from video games is much greater than previously understood. Increasing power (i.e. realism) of technology is one factor; our increased knowledge base is another. Despite some commitment to implementing our past recommendations, the industry is slipping backwards by standing still (National, 2002).

Studies (Anderson, 2000) have shown that there is a direct correlation between the consumption of violent media and aggressive behavior in children—and yet, industry-backed organizations such as the

**49**

## Effects on Search Portals

Search portals are known for having virtually everything on the Internet that people can search through. One of the big issues with the large search portals is that they rank and display websites using algorithms. They do this in order to determine where in search results a website is shown. Unfortunately, people constantly find ways to artificially "boost" their websites in the search engines and rank higher than other websites, not to mention that a search can return millions of results. Granted, while much is at your disposal, most people do not venture past the first or second page of results and may end up having to perform many searches in order to find what they are looking for.

The benefit of search roaming is that you are not trying to cater to everyone in all possible cases. Rather, you are catering to a specific audience type, and the websites that you include in your search engine are websites that fit your specific audience type and ones that you can search in real time or by using cached content (via an index on your website). Because your website is serving a specific audience type and other websites that you have included in your search engine, it is possible that your audience will utilize your website for searching rather than a search portal.

It is not likely that the major search portals will be replaced by search roaming but, rather, they will be used less frequently as Internet users find niche search roaming communities and resources that cater to specific audiences.

## Biography

Joe McCormack, an alumnus of UAT and an instructor, has been involved with web development using a wide range of programming languages on different server platforms since the mid-1990s. Joe has published two books on the subject of web-based programming and actively develops new methods and processes to automate and streamline functions to improve performance and task-handling abilities of web-based applications on different server platforms.