Editorial

Crawlers in our life

Robert J. Isaacson

The life blood of science is information. Information we create and build upon, apply, store, recall, teach and use to build the most objective development of our scientific field of interest. The Internet naturally has come to be an important part of this mission. This Utopian view is not without complications as we come to see with development of the Internet in orthodontics. Recently I was obliged to become acquainted with Web crawlers.

Web Crawlers are automated computer programs that systematically browse the Web. They survey Web sites and index the information by links tags, key words, etc. Probably every time you use a search engine, you are using the results of somebody's Web crawler. Many referrals to our site at Angle.org come from search engines and not directly by typing in our URL, http://angle.org.

You may think so what. Well these rascals recently crawled into my life and I was surprised. It seems that many firms want to have their own search engines and one way to do this is to use a Web crawler to systematically regularly survey the sites of interest on the Web and copy what they want onto their server. Why do organizations and people do this? There are very legitimate uses of Web crawlers, but there are also less constructive applications. When their server becomes the supplier of the information to a search, they can control what is displayed including what is omitted and the priority of one item over another. This raises the question, how is the priority of items displayed determined? There is not only the potential for money to be involved here, think of the censorship possibilities and how political and other bodies could control what people see who use their servers.

With this background, how does this affect orthodontic literature? If you type impacted canine into your search engine, and it goes to their server, the server determines which articles from which web sites you see and in what order they are presented to you. I always viewed this as a wonderful tool and work saver. Now I am less comfortable that I am seeing everything available without bias.

The precipitating factor here for me was when I wanted to determine the number of hits on our web site. Originally, we reported the hits as reported by Counter, the agreed official measure. These clearly were real hits, but my web site operator explained to me that in recent years the development of web crawlers has grown and now many of these hits came from Web crawlers systematically and regularly searching our site.

For the last 10 years, our long standing policy has been to make science freely available to everyone on the Internet. However, our Open Access gave free reign not only to you, but to the crawlers also. Last October we adopted a new platform with protection against Web crawlers. Now when I count our hits it is obvious that more recently activities such as crawlers were producing at least half of the hits on our web site. How much are these hits ultimately leading to proper use of the information is impossible to determine.

It's a sort of cold war with the people who would take our information for other purposes than we intended. They set up new approaches and these are countered by our people. Clearly the dichotomy between Open Access for the good of the patient and those who wants to control or generate revenue will continue to be out of sync.

This whole world is experiencing the same thing because most journals and sites want advertising revenues to support the operation of their site. Advertisers want to know what they are buying and this is determined by the traffic at your site. This has been measured by hits with research that shows some percentage of the total traffic at your site end up as users who generate revenue for the advertiser. The problem with counting hits accurately is, therefore, critical and this makes the flow of advertising dollars from the printed pages to the Internet lower and slower. We are small potatoes in this picture. Financing the sites on the Web is at the core of the paradigm shift from printed material to the digital world.