



Who we are

What we do

Our work

Talk to us

Home page / Blog / A bit of an enigma

Convert page to PDF

Press Releases

Case Studies

Microblog

Blog

## Carry on up the Internet: A bit of an enigma

**Dr. Tinkle:** It's an enigma Matron. An enigma.

**Mr. Roper:** I'm not having another one of those.

(Carry on Doctor, 1967)

Back in **Carry on up the Internet: How Big are We?** my friend added Google Analytics to the club's site. A few weeks later I was chatting to them and they said there was something odd with what they were seeing in Google Analytics; they didn't get any visitors from Google. I took a quick look at the reports and it was worse than that, they didn't get visitors from Microsoft Live Search (now Microsoft Bing), Yahoo! (soon to be Microsoft Bing), any of the search engines I recognised (soon probably all to be called Microsoft Bing, except of course, Google).

I had a quick look at their site and the Google Analytics code looked right. I tried a few searches for them on Google, Yahoo! and Microsoft Bing. Even though their name is quite distinctive they didn't come back in any of the searches.

So I did a bit more digging and found the problem; a little file on their website called robots.txt. A robots.txt file is a powerful tool for webmasters. It lets you tell Internet bots (sometimes known as web robots or simply bots) what they are allowed to look at on your site. Bots are software that runs tasks automatically on the Internet such as web spiders that are used by search engines to index the web. For example, Googlebot is used by Google to collect the information to build the index for Google's search. The robots.txt for their site denied access to all bots.

I asked them if they really meant to do that. Someone had told them about botnets and spam bots. They had an image of gangs of leather clad beastly bots rampaging around the Internet laying waste to any web site that crossed their path. Think Tron (soon to be Tron Legacy) meets Mad Max Beyond Thunderdome. After a quick bit of research they'd found a simple solution, a virtual ASBO to stop the bad bots in their tracks, using a robot.txt file to ban them from the site.

It sounds good in theory, but unfortunately it wasn't a real solution. The robots.txt file is more like the polite notice you see by the roadside in picture postcard villages asking you to drive carefully than a checkpoint that stops bots but lets other visitors in. When Googlebot and the other search engines bots visited the site they found it and obeyed the instructions in it and promptly ignored the site. Meanwhile malicious bots would simply carry on regardless.

A simple rewrite of the robots.txt file made it friendlier to good bots and let the search engines back into the site.

 [Subscribe to blog comments RSS feed](#)

### Search

via your interest

via keyword or phrase

Go

### Talk to us

Name

How would you like to be contacted?

Email  Phone

Email Address

Go

#### Site Links

[Sitemap](#)  
[Terms & Conditions](#)  
[Accessibility Statement](#)  
[Privacy Policy](#)

#### RSS Feeds

[Announcements](#)  
[Blog RSS Feed](#)  
[Twitter Feed](#)

#### Site Validation

[W3C XHTML 1.0](#)  
[W3C CSS 2.0](#)  
[W3C Accessibility WAI-AA](#)

#### Accreditation

[ISO 27001](#)  
[ISO 9001](#)  
[Institute of IT Training](#)  
[PRINCE2](#)  
[Buying Solutions](#)

#### Partners



