

14 May 2009

By: Lucian Constantin, Web News Editor

[Real-Time E-mail Harvesting on Twitter](#)

Careless users unwillingly maintain an up-to-date list of e-mail addresses

Just one day after a method of harvesting e-mails from Twitter was [exposed](#) on WebProNews, a proof-of-concept Twitter e-mail grabber was released. The technique relies on using the service's real-time search function to exploit the carelessness of users who post their addresses in status updates.

There is nothing new about the fact that spammers and e-mail marketers are using automated tools to locate e-mail addresses through Google, Yahoo! and other search engines. Some are also employing their own custom-coded robots that crawl the web specifically for this purpose. Such programs are called e-mail harvesters or grabbers and the action e-mail harvesting.

So, why would Twitter be any different in this respect? As it turns out, it is and it isn't. It is, because someone posting their e-mail address within their messages automatically makes it searchable, just like people posting it on their websites make it available on Google. This might seem obvious to many of you, but, judging by the feedback received on the issue, even knowledgeable users have overlooked this simple fact.

What makes Twitter stand apart, however, is the fact that the e-mails located through its search function are always up-to-date, as opposed to the ones located on Google, which might already be left for dead by their respective owners, exactly because of spam. This makes Twitter Search more reliable for spammers, because it orders the results by date and time, with the newest ones on top, while Google uses a page-ranking algorithm.

Additionally, Twitter Search supports operators that make it even more easier to find e-mail addresses and restrict the results to a certain period of time. For example, a query to get messages, which contain e-mail addresses, posted since yesterday, would look like: @yahoo.com OR @gmail.com OR @hotmail.com OR "email me at" OR "contact me at" since:2009-05-13. Obviously, this can be extended to include as many relevant keywords as necessary.

Then it is just a matter of automatically extracting the e-mails from the tweets. For this purpose, one can use a regular expression to identify and separate only e-mails from the rest of the text. In fact, this is exactly what a blogger has done within his newly [released](#) "Twitter e-mail grabber" script. However, despite the availability of this PoC, it's safe to assume that spammers have been harvesting e-mails from Twitter using this technique for quite a while.



Twitter Search used to harvest e-mail addresses
InterFAQs