Click here to print

Mail Online

Library of Congress attempts to archive every tweet sent since 2006 - all 170 billion of them - including the embarrassing and rude ones

By James Nye

PUBLISHED: 19:10 EST, 7 January 2013 | **UPDATED:** 19:16 EST, 7 January 2013

If like most people you don't keep track of every profound, embarrassing or even drunken tweet you post then do not worry - the Library of Congress has kindly done it for you.

The guardian of the nation's knowledge has announced it is weeks away from compiling all public tweets written since Twitter's founding in 2006, which amounts to a staggering 170 billion.

Currently the library is working how best to make the billions of individual thoughts and comments available to the public and has had about 400 requests from researchers looking to utilise the information to chart the rise of everything from the Arab Spring, to stock markets and even tracking epidemics.

The Library of Congress in Washington, DC is attempting to archive every tweet every posted since 2006

In April 2010, the Library of Congress, which was founded in 1800 signed an agreement with Twitter to get full access to all of the public tweets posted since the social networking site was founded by Jack Dorsey in 2006.

'Twitter is a new kind of collection for the Library of Congress, but an important one to its mission,' wrote Gayle Osterberg the library's director of communications in a blog post.

'As society turns to social media as a primary method of communication and creative expression, social media is supplementing, and in some cases supplanting, letters, journals, serial publications and other sources routinely collected by research libraries.'

The library has promised to keep the archive growing and currently is processing roughly 500 million tweets per day, which is significantly up from the 140 million daily messages it was sifting through in 2011 according to Osterberg.

Twitter has cooperated fully with the Library of Congress in handing over the data stream of all their tweets since they launched in 2006

Unless your account has been set to private, all public Twitter accounts and their tweets are included in the agreement.

However, the tweets included in the planned archive can only be made available to researchers six months after they are posted online.

Since the April 2010 agreement, the Library has struggled with the technological challenges that come with building and managing such a massive digital resource.

For example, the Library noted in a White Paper released on Friday 4th December that it can take up to 24 hours to perform just a single search for one term in the 133.2 terabyte Twitter archive - which they admit is not a workable option for any researchers.
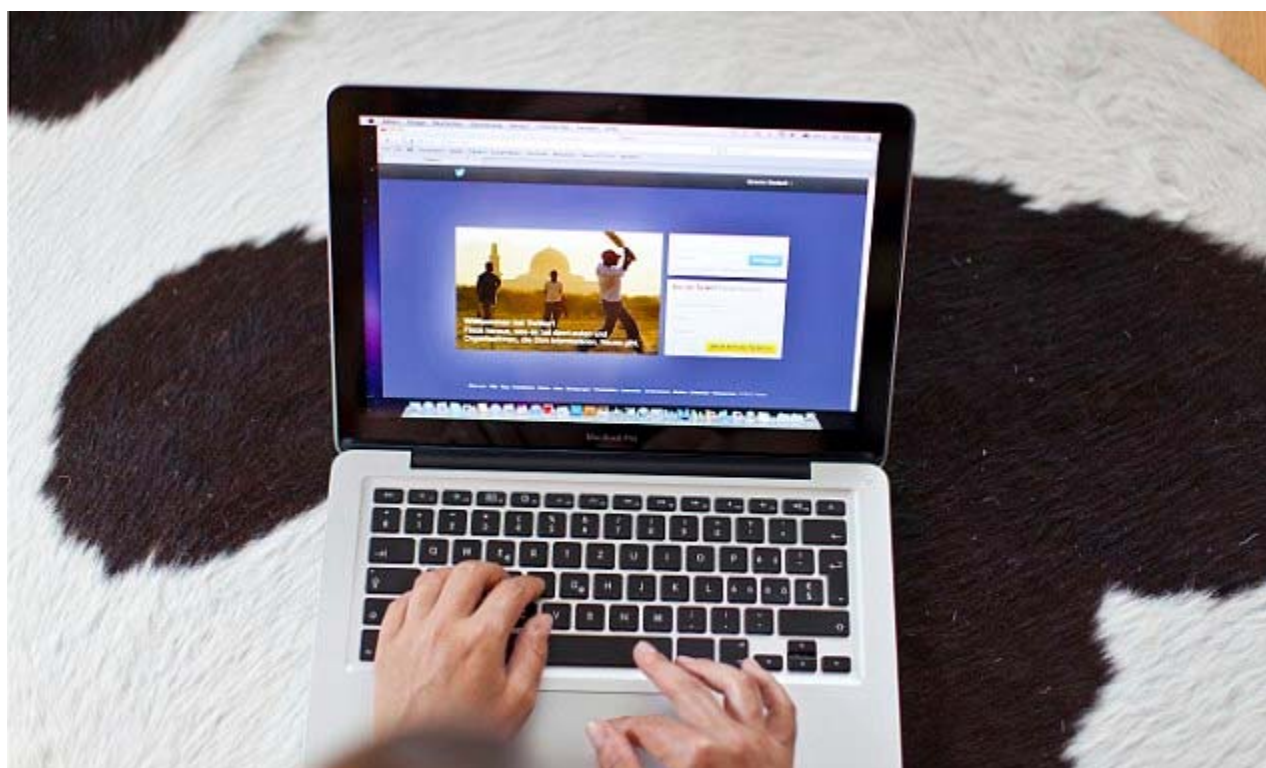
'It is clear that technology to allow for scholarship access to large data sets is not nearly as advanced as the technology for creating and distributing that data,' the Library said in the research paper.

'Even the private sector has not yet implemented cost-effective commercial solutions because of the complexity and resource requirements of such a task.'

Software architect and creator of Twitter, Jack Dorsey speaks at the Tribeca Disruptive Innovation Awards during the 2012 Tribeca Film Festival in April in New York City

Every tweet posted - including the embarrassing ones will be uploaded and available to

researchers

The intention of the project is to provide future researchers with access to today's cultural attitudes and behaviours, 'to inform scholarship, the legislative process, new works of authorship, education and other purposes.'

Expected to be finished at the end of January, the Library of Congress is still hopeful that all the teething problems will be fixed.

'This month, all those objectives will be completed,' said the white paper of the Library of Congress.

The job of archiving such a massive amount of information was contracted out the library to Gnip, a social media enterprise company based out of Colorado.

Gnip was given full access to the public tweets of Twitter's 200 million active users and has been buried deep in the process of compiling the data ever since.

'Gnip believes Twitter represents the largest archive of human behavior to have ever existed. We're thrilled that we're able to partner with the Library of Congress to help make this data available to researchers,' said a statement from the firm.

MOST READ NEWS

- Mystery as American drone is found floating off the coast of...

- 'It's cold & I'd rather die than live alone...*puts finger...

- How college students think they are more special than EVER:...

- The new cancer threat: Experts warn of coming epidemic of...

- EXCLUSIVE: Revealed, the extraordinary story of the sisters...

- Got milk? Breast pump industry booming thanks to Obamacare...

- 'Ashamed' teen who was taped laughing at horrific ordeal of...

- Lottery winner 'was fatally poisoned with cyanide ONE DAY...

- Al Roker reveals his 'bathroom accident' during White House...

- 'Liquid gold': How Tide laundry detergent is being used as...

- Controversial radio host and conspiracy theorist 'detained...

- Flu infections sweep America hospitalizing thousands and...

Comments (6)

- Newest
- Oldest
- Best rated
- Worst rated

 View all

I find it creepy

- Desi , Bel USA, 08/1/2013 06:24

Click to rate _ _ Rating   3

Report abuse

I wonder who is paying for to create this massive database and if these researchers will pay for the data requests they make? Or will this cost taxpayers a fortune to not only collect and collate the data, but also to maintain it and staff it with people to respond to requests for the data - which will then prove exactly what Lucky Moonshine said?

- Savedbygrace , Minnesota, United States, 08/1/2013 05:42

Click to rate _ _ Rating   3

Report abuse

It will be like washing the laundry -- once they are finished, there will be a new pile to complete.

- Jennifer , Knoxville-GoVols, 08/1/2013 04:16

Click to rate _ _ Rating   7

Report abuse

Seems so useless!

- tea , wash dc, 08/1/2013 04:01

Click to rate _ _ Rating   3

Report abuse

Once again, basking in the smart decision to NOT do Facebook, twitter, tumblr and whatever else is out there! I heard someone say "if you're not online, you don't exist" I'm TOTALLY fine with that!

- Emmy , USA, 08/1/2013 03:49

Click to rate _ _ Rating   13

Report abuse

Cool. 500 years from now these tweets will show how incredibly stupid and self indulgent we are. They will think we are first class narcissistic twits because we broadcast our breakfast, affairs, and bodily functions. Thank God I don't do any social media. Zero. I like flying under the radar.

- Lucky Moonshine , The South, USA, 08/1/2013 02:27

Click to rate _ _ Rating   23

Report abuse

The views expressed in the contents above are those of our users and do not necessarily reflect the views of MailOnline.

Find this story at www.dailymail.co.uk/news/article-2258752/Library-Congress-attempts-archive-tweet-sent-2006--170-billion--including-embarrassing-rude-ones.html

Published by Associated Newspapers Ltd

Part of the Daily Mail, The Mail on Sunday & Metro Media Group