

From: Bowman, Randal
To: [Shulman, Stu](#)
Cc: [Cash, Marcia](#)
Subject: Re: Stu's Email
Date: Sunday, May 28, 2017 1:33:39 PM

OK, will finish reviewing the duplicates list and wait until all the system processing is complete before doing anything more.

On Sun, May 28, 2017 at 2:31 PM, Shulman, Stu <stu@texifter.com> wrote:

Randy,

Once the final documents are in we will be redoing the groups. I think you should hold off doing more than exploring the data until the ingest is complete.

All of the features of the system are documented here:

<https://texifter.zendesk.com/hc/en-us>

Mark indicates it is the OCR process that pulls text from scanned documents that is slowing the process to a crawl on the last batch.

Stu

On Sun, May 28, 2017 at 2:25 PM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

Thank you. I am reading duplicate groups now, but have forgotten how to create codes and apply them to the groups, so will keep reading until you have time to send me directions.

On Sun, May 28, 2017 at 2:22 PM, Shulman, Stu <stu@texifter.com> wrote:

Randy,

There are a total of 29 .zip archives. 27 have completed, but the last two are still processing. This may have something to do with the presence of large attachments.

I'll keep monitoring the process.

Stu

On Sun, May 28, 2017 at 12:33 PM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

change in plans - it started raining lightly while I was typing, so no lawnmowing today. Will check every half hour or so to see if full Bears Ears data set is posted

starting about 2.

On Sun, May 28, 2017 at 9:14 AM, Shulman, Stu <stu@texifter.com> wrote:
Hold the phone.

Mark (the chief engineer) just told me there are still more comments ingesting and that those duplicate detection results were from a false start yesterday,

So, you can definitely experiment with what is showing, but we are still waiting for the final data to complete indexing and we will need to re-do the duplicate detection and clustering.

Stu

On Sun, May 28, 2017 at 8:58 AM, Shulman, Stu <stu@texifter.com> wrote:
Randy,

There were 116,417 comments altogether so the ingest and indexing takes time to run; it had to go overnight to find the duplicates. I have shared the project with you and you have full access to the data. A few notes:

- Your level of access includes the ability to delete all the data and the project.
- The duplicate detection is complete.
- The near-duplicate clustering is still running
- A search for "Bears Ears" in the raw data shows 59,225 results
- There is a "bucket" with all 59,225 you can experiment with while the clustering completes
- The biggest duplicate group has 13,372 identical comments "As a supporter of bird conservation and our public lands, I strongly urge you to protect..."

I am around tonight if you want to have a web meeting. Probably 7 PM or 9 PM would be best.

Stu

On Sat, May 27, 2017 at 6:01 PM, Bowman, Randal
<randal_bowman@ios.doi.gov> wrote:

I have registered and accessed the dashboard with no problems. However, I cannot see how to access the comments - or are they not posted yet? I will check back around 9 pm with the hope of doing a little work tonight and much more tomorrow

On Fri, May 26, 2017 at 4:17 PM, Shulman, Stu <stu@texifter.com> wrote:

--

Dr. Stuart W. Shulman
Founder and CEO, Texifter

Cell: 413-992-8513

LinkedIn: <http://www.linkedin.com/in/stuartwshulman>

--

Dr. Stuart W. Shulman

Founder and CEO, Texifter

Cell: 413-992-8513

LinkedIn: <http://www.linkedin.com/in/stuartwshulman>

--

Dr. Stuart W. Shulman

Founder and CEO, Texifter

Cell: 413-992-8513

LinkedIn: <http://www.linkedin.com/in/stuartwshulman>

--

Dr. Stuart W. Shulman

Founder and CEO, Texifter

Cell: 413-992-8513

LinkedIn: <http://www.linkedin.com/in/stuartwshulman>

--

Dr. Stuart W. Shulman

Founder and CEO, Texifter

Cell: 413-992-8513

LinkedIn: <http://www.linkedin.com/in/stuartwshulman>