

From: Shulman, Stu
To: [Bowman, Randal](#)
Cc: [Hoy, Mark](#)
Subject: Re: DiscoverText concern
Date: Friday, August 11, 2017 7:29:35 AM
Attachments: [image.png](#)
[image.png](#)

Actually thank-you Randy. We might need to send you an official DiscoverText t-shirt as a lead bug tester!

Stu

On Fri, Aug 11, 2017 at 8:06 AM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

Thank you

On Thu, Aug 10, 2017 at 11:41 PM, Hoy, Mark <mark@texifter.com> wrote:

Hey guys -

The patch for the system has been applied - the counts should all be correct now. Let me know if you see any more oddities. Thanks!

- Mark

On Thu, Aug 10, 2017 at 12:47 PM, Shulman, Stu <stu@texifter.com> wrote:

In the short term, calculating accurate numbers manually might be best, but we can explore new ways to speed up the counting process.

On Thu, Aug 10, 2017 at 11:56 AM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

Yes, that's fine. We will need total numbers supporting and opposed to the review, and - before this arose - the best idea I had for that speedwise was display 100 at a time; go quickly down each page and count how many clusters support and how many were non-codeable; then - since overwhelming majority is opposed - count how many of each item count was on the page and multiply. Example - page X had 85 clusters with opposing comments out of 100 total, 40 of those each had 8 comments (320), 45 had 7 (315)., then on to the next page. I was going to have my new assistant do that, and copy the results to a spreadsheet. If either of you have a better idea, I'd welcome it.

On Thu, Aug 10, 2017 at 11:43 AM, Shulman, Stu <stu@texifter.com> wrote:

Thanks Mark. I think we can wait until the fix is in and perhaps while you are in there, you can explore new methods of displaying some of the totals Randy might be looking for.

On Thu, Aug 10, 2017 at 11:26 AM, Hoy, Mark <mark@texifter.com> wrote:

Randy and Stu -

I see what is going on, and I have no idea why this was never noticed before. The template that controls the first list of results (when the page is initially loaded) is slightly different from the template that loads subsequent pages when using the paging controls on that page -- the counts for the initial items on the first page also include +1 for the seed file in each group - this +1 is missing from the counts on the subsequent pages. This is noticeable as well when, say, you go to page 2 and then back to page 1 - you'll notice that the counts on page 1 are off by one then...

I've got a plan for a fix and this will definitely be fixed in a patch that I can push up tonight.

In the meantime, if you need the counts now, let me know (and the details of what you need) and I can run some manual queries to get the information for you.

- Mark

On Thu, Aug 10, 2017 at 10:26 AM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

As far as I could tell, all of those actually have 2 items

On Thu, Aug 10, 2017 at 10:24 AM, Shulman, Stu <stu@texifter.com> wrote:
Definitely something odd happening I have never seen before. I have no idea why we have 1-item clusters. That should not be possible.

Inline image 1



On Thu, Aug 10, 2017 at 10:22 AM, Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

I just logged out and turned it over to Brendan, so perhaps that cluster number was a typo - I was checking at random at that point and only one of the 10-item clusters had 11. please go to some of the other numbers

On Thu, Aug 10, 2017 at 10:18 AM, Shulman, Stu <stu@texifter.com> wrote:
I see 10 in cluster 74 of June 13.

Inline image 1

On Thu, Aug 10, 2017 at 10:10 AM, Bowman, Randal
<randal_bowman@ios.doi.gov> wrote:

In trying to get ready to count the number of near duplicate comments for June 13, June 28 and and July 20, I found some apparent mistakes in the system, in that in some cases the numbers shown for items in the cluster are lower than the number that are actually in the cluster.

It appears universally that for all clusters showing 1 item - and before I looked I had wondered how 1 would even show here - actually have 2 items.

In addition, for June 13 - cluster 74 - says 10, are 11, but for clusters 75-79, 10 = 10; clusters 102-7 say 6, have 7; cluster 112 - says 5, has 6; clusters 149-181 - say 3, have 4; cluster 301 and following says 2, have 3. And for July 20, starting at cluster 181, all that say 3 items have 4, all 2's have 3, 1's have 2. This appears to be a problem only for the low-number-of-items clusters.

I had planned to count these using the numbers shown - i.e. if clusters 200-300 were all 3s, and 93 of them opposed the review, multiply to get the number of opposed, do same with support and uncodable, then go to next 100.

But if we can't rely on those numbers I am a loss on how to get to the final count. Is there something in the system that will allow us to count the numbers of comments in the lower-end clusters? Or can it be re-set to show accurately how many items are in each cluster? Only those 3 dates are involved; there were few enough near dups in the other that I have already coded them, but there are too many to code for these 3.

--

Dr. Stuart W. Shulman
Founder and CEO, Texifter
Cell: [413-992-8513](tel:413-992-8513)
LinkedIn: <http://www.linkedin.com/in/stuartwshulman>

--

Dr. Stuart W. Shulman
Founder and CEO, Texifter
Cell: [413-992-8513](tel:413-992-8513)
LinkedIn: <http://www.linkedin.com/in/stuartwshulman>

--

Dr. Stuart W. Shulman
Founder and CEO, Texifter
Cell: [413-992-8513](tel:413-992-8513)
LinkedIn: <http://www.linkedin.com/in/stuartwshulman>

--

Dr. Stuart W. Shulman
Founder and CEO, Texifter
Cell: [413-992-8513](tel: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>

Exact Duplicates

Near Duplicate Clusters

961 clusters and 52571 single items 

Items per page: 50  Page: 1 2 3 4 5 6 7 8 9 10 11 >

| <input type="checkbox"/> | Cluster | Items | Snippet |
|--------------------------|-----------------------------|-------|---|
| <input type="checkbox"/> | Cluster_251 | 2 | Grand Staircase-Escalante was designated in 1996. Since then, it has come to be known as the "Science Monument"yielding several new spec |
| <input type="checkbox"/> | Cluster_252 | 2 | I am writing to comment on the National Monument review being conducted by the Department of the Interior. Specifically, I am writing to sugge |
| <input type="checkbox"/> | Cluster_253 | 2 | I am writing in support of Browns Canyon National Monument and to comment on the Review of National Monuments currently underway. I hav |
| <input type="checkbox"/> | Cluster_254 | 2 | See attached Attachments 1522 <hr /><h3>From: DOI-2017-0002-127522-A1.pdf (OCR)</h3>/ S L2. ELAINE L. WEEKI.EY 2124 IBERIS ROA |
| <input type="checkbox"/> | Cluster_255 | 2 | America's National Monuments are national treasures, and I oppose any effort to shrink them, abolish them, or weaken their protections. These |
| <input type="checkbox"/> | Cluster_256 | 2 | Wilderness is not a luxury but a necessity of the human spirit, and as vital to our lives as water and good bread. A civilization which destroys wh |
| <input type="checkbox"/> | Cluster_257 | 2 | I am writing in support of the Katahdin Woods and Water monument. Natural resources are the strength of Our state of Maine and will be the so |
| <input type="checkbox"/> | Cluster_258 | 2 | Keep public lands in public hands! |
| <input type="checkbox"/> | Cluster_259 | 2 | Please save these monuments! Please. |
| <input type="checkbox"/> | Cluster_260 | 2 | Grand Canyon - Parashant National Monument Grand Canyon-Parashant National Monument should be permanently maintained in its entirety |
| <input type="checkbox"/> | cluster_1 | 1 | I urge you to keep the Marine National Monument Papahanaumokuakea intact. It is very important to the health of the planet and us to keep the |
| <input type="checkbox"/> | cluster_1 | 1 | I am co-founder and president of a company (LLC) that is based in Denver, CO and rooted in the outdoor recreation industry. Our mission at So |
| <input type="checkbox"/> | cluster_1 | 1 | See attached Attachments 1697 <hr /><h3>From: DOI-2017-0002-131830-A1.pdf (OCR)</h3> May 20, 2017 RECFIVFQ Gregory Tracey II ZUI |
| <input type="checkbox"/> | cluster_1 | 1 | Do not eliminate any national monuments. |
| <input type="checkbox"/> | cluster_1 | 1 | See attached Attachments 1162 <hr /><h3>From: DOI-2017-0002-106751-A1.pdf (OCR)</h3> é 'lt\$TOI€~1T\$Oii new ciatxr r Nicole Croii 225 \$ |
| <input type="checkbox"/> | cluster_1 | 1 | I am writing to you imploring that you not retract any protection of our national monuments and avoid any potential future sale of these lands. Ma |