**Search.gov Communication Draft Space**

Notes are drafted here and then transitioned to GH homebase --> <https://github.com/department-of-veterans-affairs/va.gov-team/blob/master/products/on-site-search/meeting-notes/search.gov/search-gov-meeting-notes.md>

**Aug 11th Follow up Email regarding VA Videos**

**Hi Amani,**

Thanks for meeting with us last week I know John was pleased to connect with everyone one last time.

We wanted to follow up on the Video discussion that we had during that call. We have identified a few behaviors (see below) that aren’t what we anticipated and we are hoping you could look into for us. We also wanted to get a better sense of if the Video API is available for API clients as we are exploring this as a potential improvement for va.gov content.

Lastly, please note that I will be out of the office for the next two weeks Aug 16-27 so we may be slow to respond during that time.

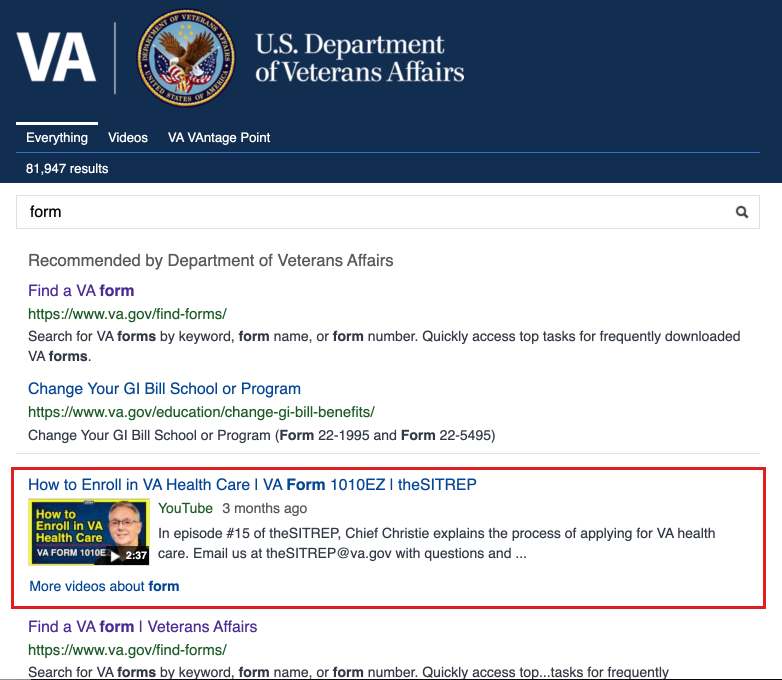
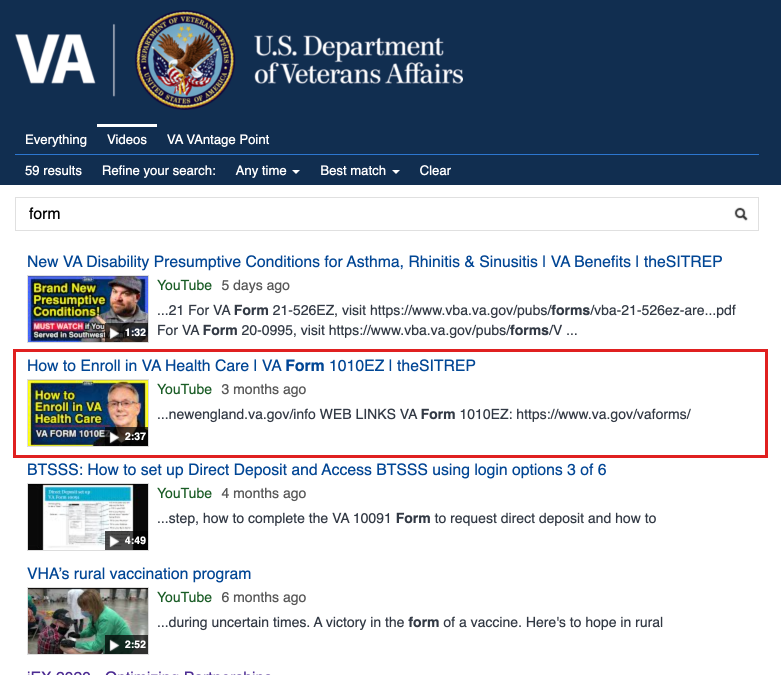
Thanks,

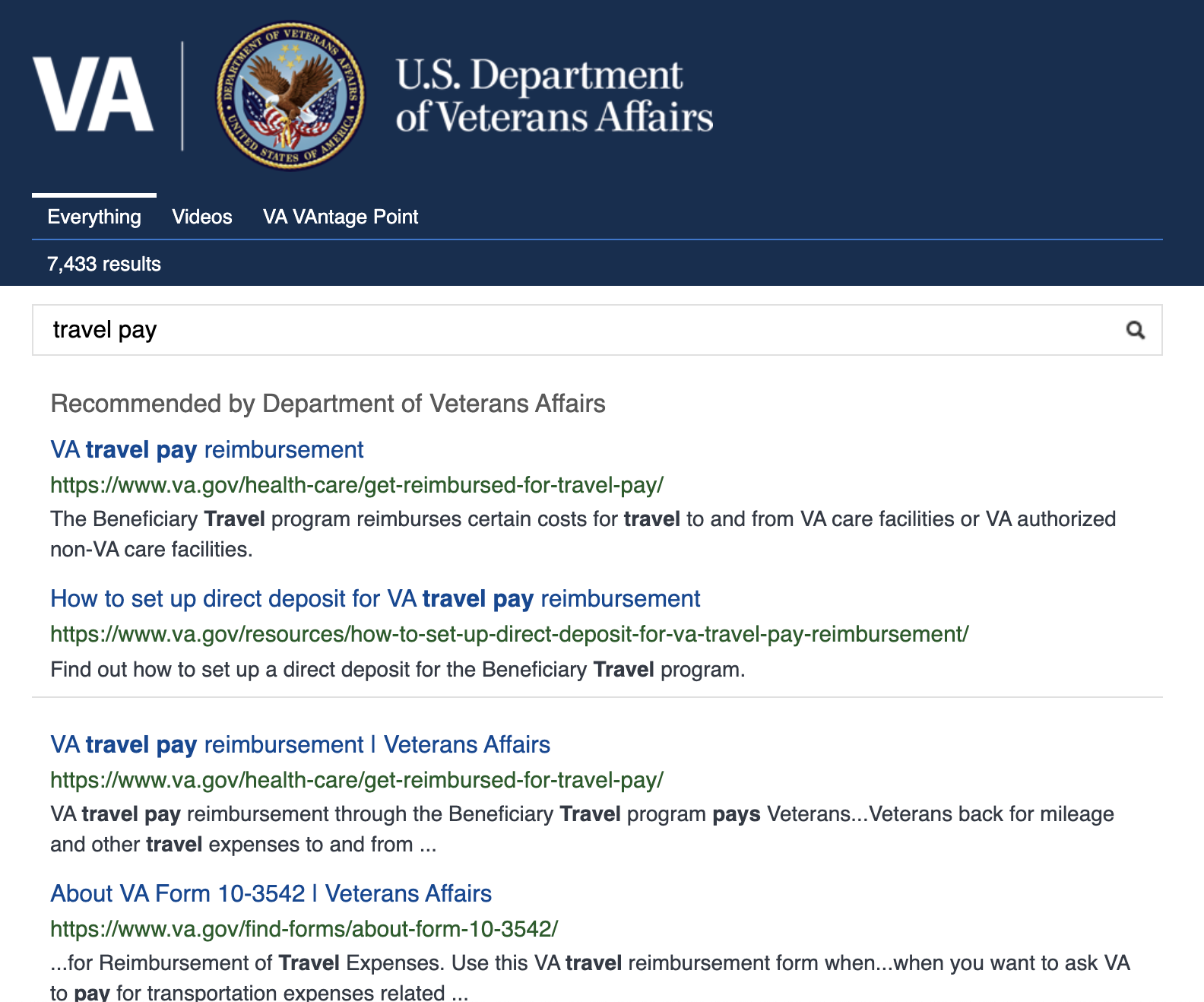
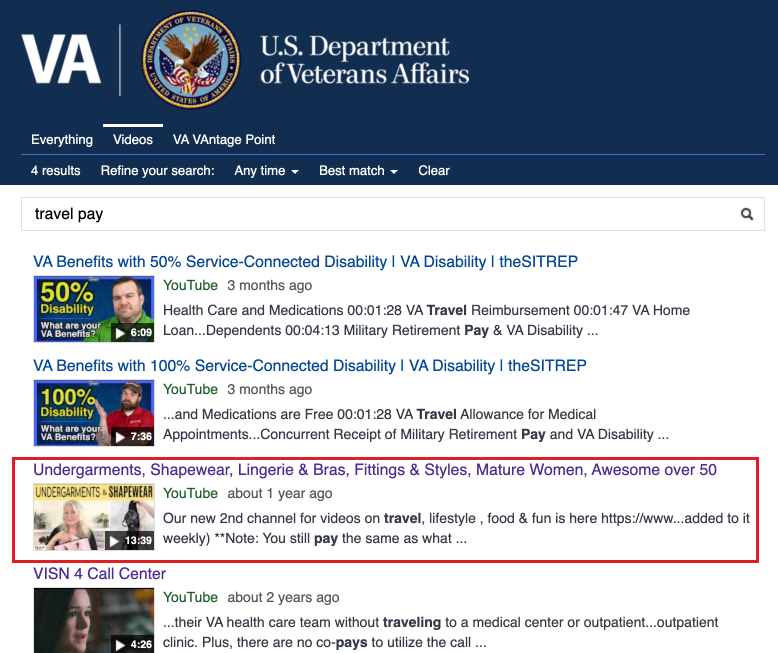
Denise

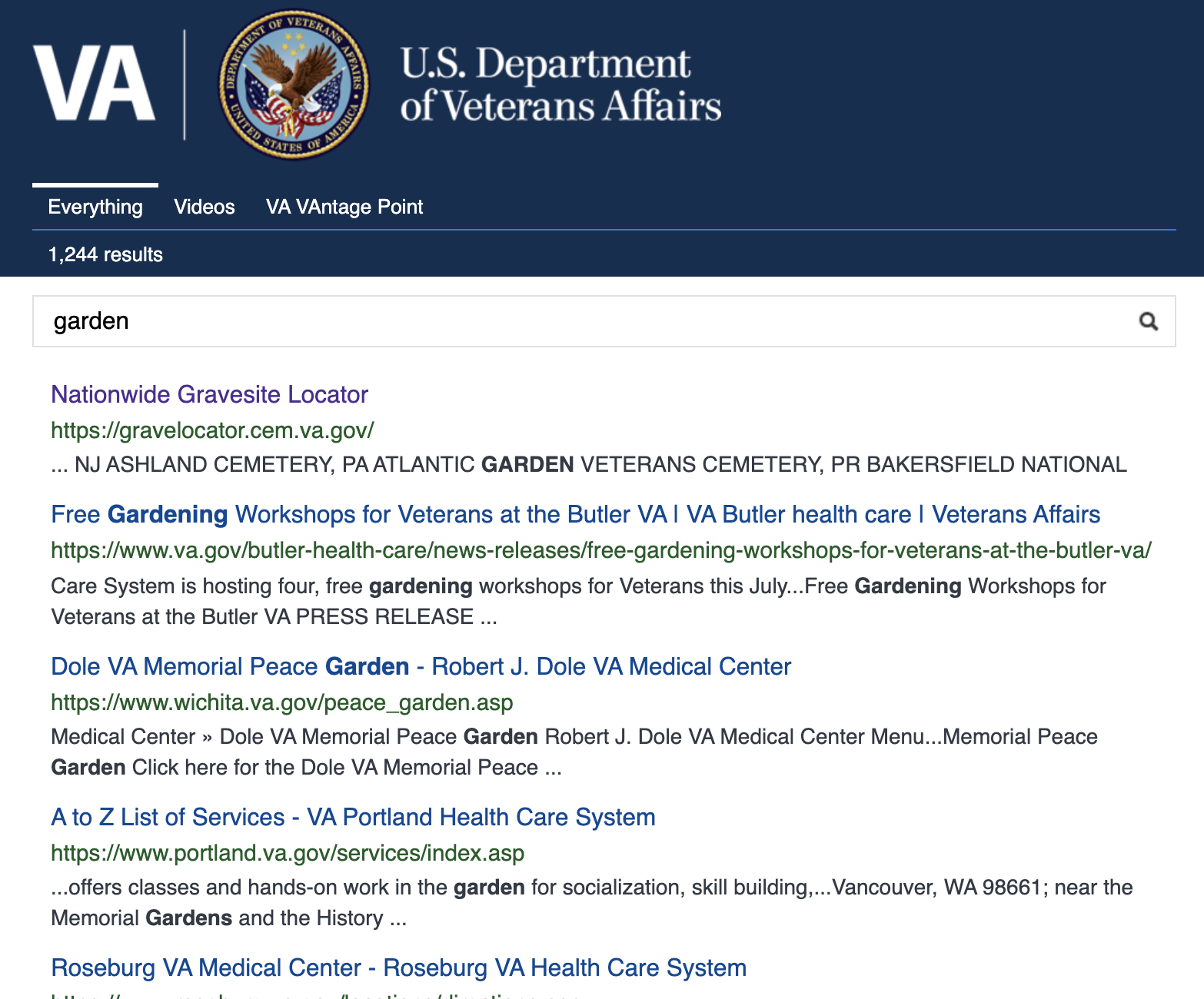
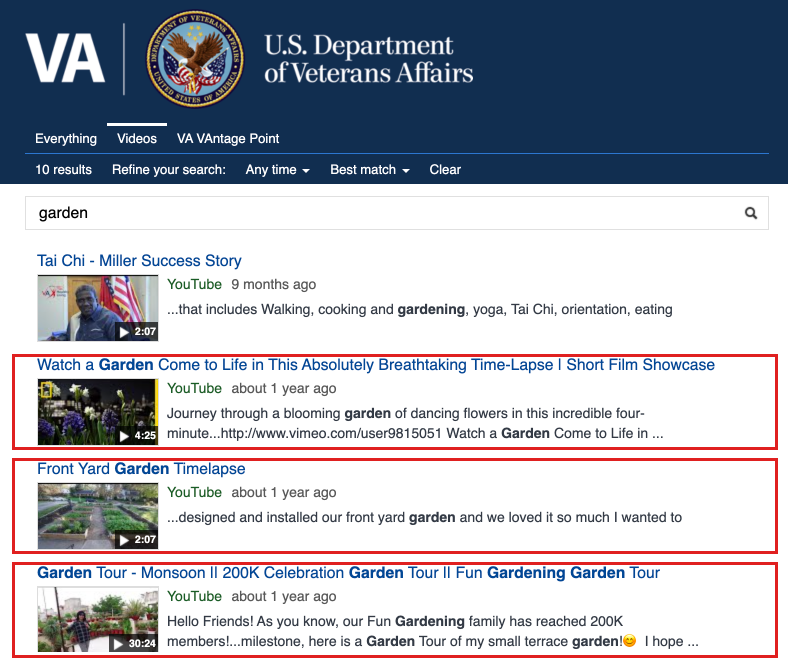
Here is the info about the potential bugs/weird behaviors that we identified.

The following bugs were found on the Search.gov preview in regards to video:

* One video result shown on "Everything" results does not always match first video result on "Videos" tab.
* One video result not always showing on "Everything" results tab even when there are video results.
* Some search terms pull up video results from random YouTube channels that we have not added on Search.gov's admin.

Search term: "**form**"  
  


Search term: "**travel pay**"  
  


Search term: "**garden**"  
  


**{Denise will be moving these notes to GH}** [**https://github.com/department-of-veterans-affairs/va.gov-team/blob/master/products/on-site-search/meeting-notes/search.gov/search-gov-meeting-notes.md**](https://github.com/department-of-veterans-affairs/va.gov-team/blob/master/products/on-site-search/meeting-notes/search.gov/search-gov-meeting-notes.md)

**June 15 - Indexing & Relevance Meeting**

Purpose:

* To better understand how content is being indexed and what we can alter to impact the relevancy score.

Indexing + Relevance Discussion:

* Resources and Support content is simply not surfacing in Search -- even when you search the H1 or an article. Take *Can I be buried in Arlington National Cemetery?* as an example. When we meet with Amani this is going to be one of the most pointed Use Cases we will surface. Is this an "us" thing -- we just need to tweak our indexing model -- or a "them" thing? (Note what [UK.gov does](https://insidegovuk.blog.gov.uk/2014/06/13/how-gov-uk-site-search-works/) -- *Results from the* [*‘mainstream’ services and information*](https://insidegovuk.blog.gov.uk/2014/06/02/what-is-mainstream-content/) *have a higher weighting than ‘departments and policy’ content, to make sure that the general public can easily find what they need without being confused by professional or specialist information.)*
* The Articles for R&S are found here: https://www.va.gov/resources/search/articles.json , if that helps us diagnose why they are not being indexed.
  + Idea: Can we boost modern VA.gov pages vs. legacy? How would this be done (via source location or artificially boosting the popularity of those pages?)
* What’s the latest on Collections?
* We talked to ElasticSearch about their new App Search product -- <https://www.elastic.co/app-search/> -- and they mentioned a meeting with you/Dawn. Any update coming out of that?

Other (if time allows)

* Able to ***remove stop words***?
  + *Elasticsearch by default*: a, an, and, are, as, at, be, but, by, for, if, in, into, is, it, no, not, of, on, or, such, that, the, their, then, there, these, they, this, to, was, will, with
  + *Additional stop words:* about, am, any, become, became, can, could, did, do, does, doing, get, getting, had, has, have, how, how's, I, I'm, it's, its, just, my, myself, so, than, that's, them, too, we, who, what, what's, when, where, which, who, who's, whose, why, would, yet, you
* Feasibility of creating an ***API for best bets*** in the future?
* Any way to run a “***best bet spreadsheet”*** report? Could we extend the spreadsheet to include other fields suchs as categories for our own purposes?
* Could we add a ***filter to click tracking data*** to sort by data range?

June 17th Notes:

* They want to validate the stop words then will come back to us
* They only use sitemap to crawl our content
  + They manually crawl subdomains (but admittedly not well)
  + Not up to their standards but up to their capacity.
* They have been exploring the idea of letting us boost on our own.
* DESCRIPTION/SNIPPET content - quality of meta descriptions was spotty at best and - can do that at a domain level.
* Dawn doesnt think that the elastic search app will work very well in our context.
  + VA proxies requests and therefore the location feature wouldn’t work.
* VA.gov is not mandated to use search.gov
  + Search.gov is free
  + She thinks it would be very expensive for us.
  + Before 7500$/mth - 90K a year
* Amani used to work at yext.

Action Items for **Search & Discovery**

* + Add R+S content to the sitemap
  + Determine if any other content is not showing up on the sitemap
  + To boost the modern pages we need to identify if the entire WWW is considered modern or are there certain domain folders. Then notify search.gov who can use the booster flip to promote modern content

Action Items for **Search.gov**

* + Continue exploring use of stop words
  + Run a crawl to seed the index with R+S articles <https://www.va.gov/resources/search/articles.json>

<https://www.va.gov/resources/>

* + Share GH Repo of what is being indexed (already shared - <https://github.com/GSA/search-gov-agency-collaboration/tree/master/va>
  + Place logic at domain level to enable metadata descriptions to be enabled for all modernized va content

**June 4 - email response**

Thanks for your patience! See my responses below in **bold**, let me know if any of the formatting makes it hard to read:

1- Low CTR Terms

We noticed a few puzzling things when we look at the list of ‘Low Click Thru’ terms each month. Many have a 0% click thru despite the fact they have been optimized with best bets. A few examples include “health topics” “ebenefits” “travel pay” all which have associated results in best bets.

We dug a little deeper using “health topics” as an example and we discovered the following:

· “Health topics” is not a top query in google analytics

· It has a 0% click thru rate but is also a top query.

· We do have two best bets for that query which are a direct match.

· Health topics is one of the modules.

Something fishy is certainly going on hoping you can help pinpoint what that may be. We are wondering if it is a data issue or maybe we are misinterpreting the information. Any thoughts that may help us understand what is happening?

**For Health Topics, it looks like we’re getting hit with that query around 144 times a day, which makes me think it might be more of a “health check” originating from your system to make sure we’re up and search is working. Is there anything implemented on your end that would be sending this consistent query? The user agent claims to be “Vets.gov Agent” for these requests. In general, it does look like you're calling our Clicks API successfully.**

2- Hardcoded terms on Algorithm List

On the suggestion database list, you sent on May 20th we noticed that there are a few terms marked “false” that are actually on the hardcoded list. A few examples include 'travel pay', ' mental health', 'home loans', ' loans', 'disability benefits'. Can you tell us if terms on the hardcoded list are used to filter out any terms that may appear via the algorithm? Or should we expect duplicate terms may appear? (similar question below)

**This is on me - we don’t have an easy way to export all suggestions directly from our database, which has a “protected” flag to indicate the hardcoded entries. I pulled this list from our data visualization tool which is missing that key field, and looks like I did the lookup improperly. Attached is a corrected list.** (added to our folder - <https://github.com/department-of-veterans-affairs/va.gov-team/blob/master/products/on-site-search/meeting-notes/search.gov/VA%20Typeahead%20Suggestions%206.1.2021.xlsx>)

3- Validation how 5 suggestions are selected. **(Corrections/Comments are in bold)**

As we monitor how typeahead is working we wanted to validate our understanding of how the 5 suggestions that are presented to users are selected. That led us to review our "how it works documentation and thought it would be good to get confirmation from you. Would you be able to review the following description of “how it works” and let us know if there are any discrepancies or further details, we can add?” One point in particular we aren't sure of is "Remove any hardcoded suggestions from the algorithm list" can you tell us if the hardcoded list is used to filter out anything that may come from the algorithm as a duplicate (this may be related to topic [#2](https://searchusa.zendesk.com/agent/tickets/2)) but we aren't sure.

“How Typeahead Works”

· Up to 5 suggestions are presented to a user after they have typed 3 characters or more in the query input box. **Confirmed**

· Suggestions are derived using two methods:

1. Search Algorithm that leverages the searches performed on VA.gov website.

· A given term/phrase must be searched 5 times in a 24-hour period to appear as a suggestion. **A phrase must now be searched 30 times in a 24 hour period to appear.**

· If a suggestion does not meet those requirements, then it will be removed from circulation after a 30-day period. **Yes.**

· Certain queries will not appear as suggestions for privacy and security purposes. (e.g., Names, SIN, etc.) **We filter out SIN via regex, but we don’t have a systematic way to filter names.**

· Terms related to geographical locations will not appear including cities and states. **We are only filtering out state names and (most) state abbreviations.**

1. A hardcoded list of specific terms curated to align with VA topics. Terms that are added to this list include:

· hard to spell terms, particularly long ones (benefits, government)

· alphanumeric form titles (10-10ez)

· common veteran tasks

· top [va.gov](http://va.gov/) (<https://gcc02.safelinks.protection.outlook.com/?url=http%3A%2F%2Fva.gov%2F&data=04%7C01%7C%7Cf107e62060f149f5e2e408d9206e5128%7Ce95f1b23abaf45ee821db7ab251ab3bf%7C0%7C0%7C637576480459196473%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=fv4GTOT5l5ins87HueR85Zx4cFHopolS11tsND9C%2FdY%3D&reserved=0>) searches (benefits)

· best bets (e.g., Covid-19 vaccine)

· promoting va content to aid in wayfinding (e.g., Find a form or how do I?)

· The 5 suggestions that are presented to users are selected from the suggestion database (combination of both the algorithm terms and the hardcoded terms).

· The suggestions are ranked using the following rules:

\*\* Use query to find ALL possible matches

\*\* Sorted list by hardcoded/canned suggestion

\*\* Remove any hardcoded suggestions from the algorithm list

\*\* Then by algorithm / non-canned suggestion

\*\* Sort by best match (still preserving the above order)

\*\* Take the top 5

**The unique “key” in our index is the phrase itself, so we don’t store duplicates based off of source (hardcoded, best bets, or algorithmically determined). The order is determined by popularity - best bet keywords and the hardcoded ones have an artificially high popularity set, so these will typically return first in the list. If popularity is the same, they are listed in alphabetical order as a tiebreaker.**

Search.gov's implementation of type-ahead does support fuzzy string matching to help with typos. **No, we match on the characters entered in the search box (once there are three or more characters entered) - fuzzy matching only goes into effect after a user searches.**

Let me know if there are any further questions after you get a chance to review!

**May 26 - email sent to search.gov**

Hi Amani,

Hope you are doing well and looking forward to the long weekend. We wanted to let you know typeahead is now live to all VA.gov users and overall we are seeing a high CTR for our suggestions. We also wanted to reach out to ask if you could help us interpret/ validate a few things. Hopefully

1- Low CTR Terms

We noticed a few puzzling things when we look at the list of ‘Low Click Thru’ terms each month. Many have a 0% click thru despite the fact they have been optimized with best bets. A few examples include “health topics” “ebenefits” “travel pay” all which have associated results in best bets.

We dug a little deeper using “health topics” as an example and we discovered the following:

· “Health topics” is not a top query in google analytics

· It has a 0% click thru rate but is also a top query.

· We do have two best bets for that query which are a direct match.

· Health topics is one of the modules.

Something fishy is certainly going on hoping you can help pinpoint what that may be. We are wondering if it is a data issue or maybe we are misinterpreting the information. Any thoughts that may help us understand what is happening?

2- Hardcoded terms on Algorithm List

On the suggestion database list, you sent on May 20th we noticed that there are a few terms marked “false” that are actually on the hardcoded list. A few examples include 'travel pay', ' mental health', 'home loans', ' loans', 'disability benefits'. Can you tell us if terms on the hardcoded list are used to filter out any terms that may appear via the algorithm? Or should we expect duplicate terms may appear? (similar question below)

3- Validation how 5 suggestions are selected

As we monitor how typeahead is working we wanted to validate our understanding of how the 5 suggestions that are presented to users are selected. That led us to review our "how it works documentation and thought it would be good to get confirmation from you. Would you be able to review the following description of “how it works” and let us know if there are any discrepancies or further details, we can add?” One point in particular we aren't sure of is "Remove any hardcoded suggestions from the algorithm list" can you tell us if the hardcoded list is used to filter out anything that may come from the algorithm as a duplicate (this may be related to topic #2) but we aren't sure.

*“How Typeahead Works”*

* Up to 5 suggestions are presented to a user after they have typed 3 characters or more in the query input box.
* Suggestions are derived using two methods:

1. Search Algorithm that leverages the searches performed on VA.gov website.
   * A given term/phrase must be searched 5 times in a 24-hour period to appear as a suggestion.
   * If a suggestion does not meet those requirements, then it will be removed from circulation after a 30-day period.
   * Certain queries will not appear as suggestions for privacy and security purposes. (e.g., Names, SIN, etc.)
   * Terms related to geographical locations will not appear including cities and states.
2. A hardcoded list of specific terms curated to align with VA topics. Terms that are added to this list include:
   * hard to spell terms, particularly long ones (benefits, government)
   * alphanumeric form titles (10-10ez)
   * common veteran tasks
   * top va.gov searches (benefits)
   * best bets (e.g., Covid-19 vaccine)
   * promoting va content to aid in wayfinding (e.g., Find a form or how do I?)

* The 5 suggestions that are presented to users are selected from the suggestion database (combination of both the algorithm terms and the hardcoded terms).
* The suggestions are ranked using the following rules:
  + Use query to find ALL possible matches
  + Sorted list by hardcoded/canned suggestion
  + Remove any hardcoded suggestions from the algorithm list
  + Then by algorithm / non-canned suggestion
  + Sort by best match (still preserving the above order)
  + Take the top 5
* Search.gov's implementation of type-ahead does support fuzzy string matching to help with typos.

**April 21- Received response Email**

Hi Denise,

I'm so sorry if this wasn't clear after our last call, but the metadata descriptions isn't a feature we currently have. We do have the meta descriptions indexed for content matching, but not a built-in toggle to flip from the snippet to meta descriptions in the SERP display.

We've started scoping out this feature in the past, but haven't had much interest from our customer base yet. One thing we discussed during our last call was understanding your relative priority of the features we've talked about - a list is below:

* Boosting / Weighting Domains in Search Results Ranking
* Handling Synonyms
* Handling Homophones
* Metadata Descriptions Option
* Increased Control over Search Suggestions

That'll definitely help us as resources become available to work on some of these features.

Again, apologies for the miscommunication here!

Best,

Amani

**[April 21] Email sent Re: HTML metadata descriptions**

Hi Amani,

As you know from our last discussion, we have been reviewing the html descriptions and believe that by enabling them in the UI we would be improving the overall veteran experience. We would like to ask a few clarifying questions before we move forward.

* ***Available Descriptions:*** From our assessment ([see attached spreadsheet](https://docs.google.com/spreadsheets/d/1uApEPZ5gY3pkLMIzsos0koz64Z7-Vc4-8DerFREKac0/edit?usp=sharing)) we believe that most of our content has metadata descriptions (defined within the html of each page) that are more descriptive than what appears today. If we turn the feature on, we believe these descriptions will appear instead of what is there today. Can you confirm that?
* ***Unavailable Descriptions:*** We do know that some content (eg ebenefits content) does not have an html description and we believe that the description would not change from what is present today. Can you confirm that?
* ***Turning Descriptions ON:*** How easy is it for you to “flip the switch” to turn the html metadata descriptions on for va.gov? What would the turnaround time look like once we give you the go ahead?
* ***Turning Descriptions OFF:*** In the event that the descriptions don’t improve the experience or things look out of place, how easy is it for you to reverse our decision and return to our current state? (ie. flip the switch again to turn them off)
* ***Other:*** Do you foresee any other challenges or obstacles that we should be aware of?

Thank you for clarifying our understanding regarding these descriptions.

**[April 19 Received] Search.gov Response Email**

Hi Denise,

Thanks for your patience - our answers are below!

**Click Tracking**

* What is the average click tracking rate for other clients?
  + The average is between 50-75% over the last 7 days for those tracking click data. Not all of our partners have implemented click tracking, so our system-wide metrics (that you'll see in the Analytics overview screen) are much lower than what we see on a site-by-site basis.
* Have you identify baseline/benchmark/industry standard CTRs
  + This is fairly content dependent, and across boutique search providers there aren't published "standards" per se, but we do think the Search.gov benchmark listed above is a good goal. It looks like VA is getting around 40% right now, which is good to see!

**Search.gov Roadmap + Features**

* Do you have a list/info about all [search.gov](http://search.gov/) features that are available to us? (routed queries, best bets, similar content etc?)
  + The links below will be helpful to you!
    - [*https://search.gov/manual/*](https://search.gov/manual/)
    - [*https://search.gov/manual/site-launch-guide.html#add-features*](https://search.gov/manual/site-launch-guide.html#add-features)
* Do they have a list of other government sites that use [search.gov](http://search.gov/) exclusively? We would like to explore their experiences and review how they implement various features.
  + Here are some example sites - the hosted SERPs being our out-of-the-box solution, and the Results API examples being similar to your set-up. All of these are using the [search.gov](http://search.gov/) index (not Bing).
  + List of Sites:
    - Hosted SERP:
      * <https://home.treasury.gov/>
      * <https://www.uscis.gov/>
      * <https://www.nps.gov/index.htm>
      * <https://www.uspto.gov/>
    - Results API:
      * <https://www.tsp.gov/>
      * <https://studentaid.gov/>

**Best Bets**

* What is the total number of best bets we can create? We believe it is something like 391 but we are wondering if it is something that we could increase over time?
  + There isn’t an imposed limit - is the team running into issues adding additional best bets? We have some sites with over a thousand best bets, so you should be all clear to add as many as you'd like.
* We are also getting ready to implement click tracking for our best bets and wanted to ensure we understand where to find that information once it is connected. We believe it will appear here but wanted to clarify. (see attached screenshot - blue circle)
  + Yes, that is where the analytics will appear once implemented to add the module codes.

**Analytics Questions**

The table in the screenshot also posed a few questions that we wanted to clarify.

* What are the different “modules”?
  + The modules are the different components of your search experience - more information on each can be found on our [search.gov](http://search.gov/) site, under the Search Features.
    - <https://search.gov/manual/module-codes.html>
    - <https://search.gov/manual/site-launch-guide.html#add-features>
* How are the impressions calculated?
  + We count an impression anytime that module appears in the returned search results*.*
* Why is there an “Average CTR” for the modules that don’t have any Clicks?
  + These are system-wide averages across sites that use Search.gov, not metrics for your specific site.

Let me know if there are any follow-up questions, and have a lovely weekend!

**Preparing Email [ sent on April 9th- expecting reply by April 16th]**

*Click Tracking*

* What is the average click tracking rate for other clients?
* Have you identify baseline/benchmark/industry standard CTRs

*Search.gov Roadmap + Features*

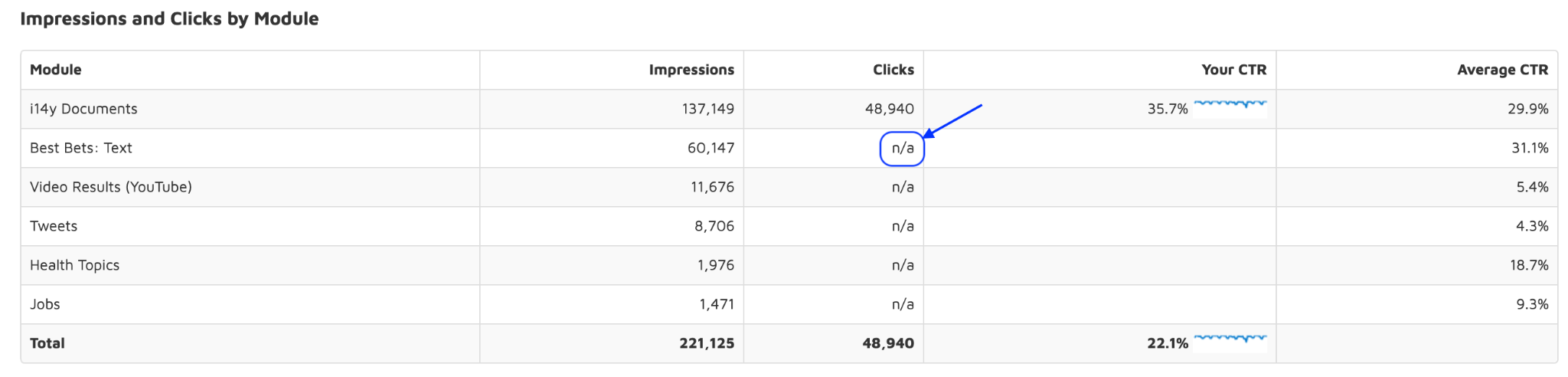
* Do you have a list/info about all search.gov features that are available to us? (routed queries, best bets, similar content etc?)
* Do they have a list of other government sites that use search.gov exclusively? We would like to explore their experiences and review how they implement various features.

*Best Bets*

* What is the total number of best bets we can create? We believe it is something like 391 but we are wondering if it is something that we could increase over time?
* *We are also getting ready to implement click tracking for our best bets and wanted to ensure we understand where to find that information once it is connected. We believe it will appear here but wanted to clarify. (see attached screenshot - blue circle)*

*Impressions and Clicks by Module*

* *The table in the screenshot also posed a few questions that we wanted to clarify.*
  + *What are the different “modules”?*
  + *How are the impressions calculated?*
  + *Why is there an “Average CTR” for the modules that don’t have any clicks?*

**

**Preparing for Meeting on Monday, April 5th.**

**Agenda**

1. Logic Levers - Threshold + Deprecate
2. Hardcoded List of Typeahead Suggestions
3. Result Relevancy Questions
4. Metadata Descriptions
5. ~~Analytics Variations~~

**1- Logic Levers**

Discuss the details. We are currently thinking for our purposes a threshold of 40 identical queries/day to deprecate after 15 days would make sense, but we would like to discuss how Search.gov reached the proposed 30 identical queries/day to deprecate after 5-7 days so we understanding your thinking and can find the right balance.

*Related Questions:*

* Why is search.gov trying to improve the feeling of recency in government space?
* When does search.gov expect to make adjustments? Can search.gov keep us informed with any changes they make?
* How easy is it to change these going forward? A few adjustments may be needed until we find the right balance.f

**2- Typeahead Suggestions**

I have attached the list we have so far. We have an additional 100 terms which we would like John to add a best bet for before we include them.

*Related Questions:*

* Could they add them by April 7th?
* No changes between **April 12-25th**: We plan to run a POC during this time. Can we request that no changes are made during that time so we can analyze the experience and make adjustments based on findings?
* Can search.gov clarify how suggestions are matched (alphabetically/popularity)
* Going forward we anticipate we will identify new terms, discuss frequency of adding new terms to the hardcoded list - *monthly*?

**3- Result Relevancy**

We would like to improve the relevancy of results by implementing search best practices - [more info on Nielsen Norman Group](https://www.nngroup.com/articles/internal-website-search/)Is search.gov using these today?

*Related Questions:*

* Remove 'stop words' (Example: remove sentence structure words that reduce result quality ('to', 'get') searching for “how to get prescription” doesn’t produce the results we want however searching for “prescription” shows better results) [Elastic search + stop words](https://qbox.io/blog/how-to-use-elasticsearch-remove-stopwords-from-query/)
* Stemming "Stemming refers to removing endings (like suffixes) from words, and reducing them to their morphological ‘stem.’ ... Although stemming will boost recall, it can also lower precision by returning results that are not relevant, but happen to include the stem of a query word." [Elastic search stemming](https://www.elastic.co/guide/en/elasticsearch/reference/current/stemming.html)

**4- Metadata Descriptions**

We are evaluating what the experience with the metadata descriptions turned on would be like and would like to discuss

*Related Questions:*

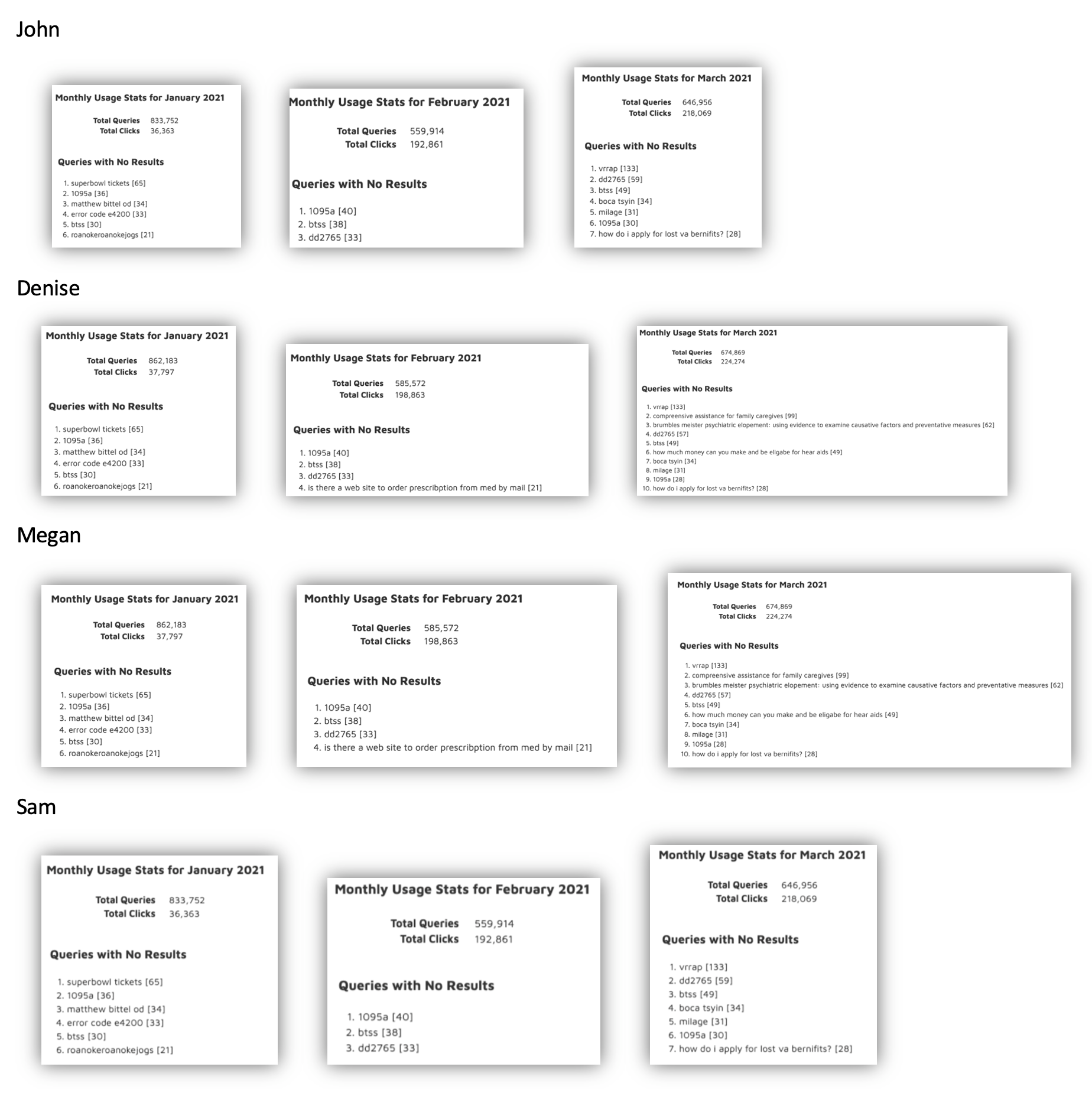
* How are the descriptions that are there today being generated?
* eBenefits content does not have any descriptions - will they remain blank or what will be pulled in?
* What do you need from us to turn on the metadata descriptions if we decide we want to do that?
* What is the current character limit? We understand that we would need to limit it further if we wanted it any shorter.

**5 - Analytics Variations.**

We have discovered that the team is seeing and reporting on different data sets, can you clarify. John/Sam are seeing one set while Denise/Megan are seeing another (see below).

*Related Questions:*

* *How do we rectify?*

****

**Search.gov Meeting Notes:**

Stop words = enabled

More can be added

20 words

Stop words:

a an and are as at be but by for if in into is it

no not of on or such that the their then there these

they this to was will with

Stemming is enabled

Promote tag is available

Synonyms - high on the list

<https://search.gov/manual/metadata.html>

<https://search.gov/manual/what-searchgov-indexes.html>

<https://search.gov/manual/ranking-factors.html>

**Friday Feb 19 - Email to Search.gov** *Topics with details for each*

**Blacklisting**

* To aid in communication and expectations we would like to propose the following schedule. Dates can of course be edited but we wanted to get a sense of timing on both our end and theirs. (see implementation plan chart on [GH#19763](https://github.com/department-of-veterans-affairs/va.gov-team/issues/19763))

**Clarify US.gov implementation**

* When we asked for a US government site that is also powered by search.gov they shared the us.gov site however it appears this is powered by Bing. Is there a joint implementation of some kind or what is the implementation? Is that something we may be able to leverage?

**Removing State references from suggestions**

* When do they plan to have state references removed?
* Can you confirm that you do not plan to remove the acronym for Virginia “va” since we are concerned that will impact the experience?

**Friday Feb 19 - Meeting with Search.gov**

Agenda:

**Spelling Suggestions**

* Validate our understanding of when fuzzy matching will be applied? Is it only if no results are found? Show journey map on Miro.
* Can we address the terms that we notice are not spelled correctly but are presenting results (eg. goverment and Benfits?)
* Provided example (<https://search.usa.gov>) - when searching for goverment users see results for government but can then add results for the misspelling.

**Blacklisting**

* Discuss sample list of terms to beremoved and misspellings

**Routed Queries**

* No blockers, we haven’t connected it yet but don’t anticipate any concerns.

**Search descriptions/Metadata [If there is time]**

* Is it possible to limit the number of characters on search result descriptions?
* It doesn’t seem like the metadata description for each page is appearing as a search result description, can this be fixed?

Use page description.

* Filter it out completely - (will it cause harm?)
* Any city names wont be pulled out
* 3 or 4 lines show in snippet - truncation
* We will be added to “user our page descriptions” at the domain level because not all have the same level of data

Type-ahead Lists

**• Go ahead and send us your full lists (for adds, removes, and misspellings) in priority order. We'll do these in batches weekly (as it's a bit manual on our end), likely aiming to cover 50 or so per list per week.**

Spell Correction Questions

• If a user searches a misspelled word, will the user see the results for the correct spelling and see “search instead for (misspelled word)”? Or will they see search results for the misspelled word and see “did you mean (correctly spelled word)”?

**• The API will return the property "spelling\_correction" in the response with the corrected term and the results will reflect the corrected term. The way this is represented to the user is up to your implementation!**

• Will a misspelled search term with any number of results still trigger the web:spelling\_correction feature (e.g. benfits (https://www.va.gov/search/?query=benfits) )

**• Yes - if there are documents in the index that match that misspelled term, those will be returned in the results, ahead of any fuzzy matching being applied, and the "spelling\_correction" will return as null in the API response.**

• How are misspellings determined? Is this changeable?

• **The corrections are determined by Elastic's fuzzy matching and are unfortunately not configureable per search experience. That said, there are likely improvements we can make to this fuzzy matching in the future - please pass over any specific examples / wishlist items and we'd love to add them to track as a customer request for prioritization on our roadmap.**

• Is there a government agency using this Search.gov feature you can point us to see an example?

• **The spellcheck / fuzzy matching is built-in to all search experiences (and impacts search results), but the representation in the SERP depends on how a customer implements it, if they're using the API. In the out of the box experience, it's represented as "Showing results for [[corrected term]]" - here's an example: https://search.usa.gov/search?utf8=%E2%9C%93&affiliate=usagov&query=resourcs**

Let me know if this helps give context for the UX of the experience! Please reach out with any additional questions.

**Bing**How is <https://search.usa.gov/>set up? Search.USA.gov was sent as an example of Search.gov’s spell check but it was later stated they were using Bing for spell check. Is <https://search.usa.gov/>utilizing both Bing and Search.gov?

**Filtering**

How will filtering work in the future? Is there anything we need to do to prepare content for this?

**February 12, 2021 - Follow up Questions for search.gov team**

Note: **Set meeting about Routed Query issues with Amani and Sam next Friday 3-4pm**

* **We plan to start by reviewing the suggestions for the top 50 search terms from 2020**? That is a list of XXXXXX terms for us to review. Given there is manual work here for your team to complete, do you have a limit to the refinements you can manage?

E.g.,

* **Can you provide further explanation on how the** [**web:spelling\_correction**](https://open.gsa.gov/api/searchgov-results/#webspelling_correction) **spelling feature works?   
    
  If a user searches a misspelled word, will the user see the results for the corrected spelling and see “search instead for ‘misspelled word’”? Or, will they see search results for the misspelled word and see “did you mean ‘correctly spelled word’”?  
    
  How are misspellings determined? And is this changeable?  
    
  Will a misspelled search term with any number of results still trigger the web:spelling\_correction feature? (Ex:** [**benfits**](https://www.va.gov/search/?query=benfits)**)**

**Is there a government agency website that is using this -- and we can see how it works?**

February 10, 2021 - Follow up Questions for search.gov team

* **Can we add terms to the typeahead suggestion list**?

Could we integrate a list of terms that will improve users ability to access content?

*Examples:*

* 1. ‘Covid-19 FAQs’ (currently if you search for covid you aren’t presented with a path the the FAQs)
  2. terms to expand on a certain query such as ‘ebenefits your records’, ‘ebenefits direct deposit’, ‘ebenefits disability compensation’, and ‘ebenefits update contact information’
* **Can we remove terms from the typeahead suggestion list?**

*Examples:*

* 1. Locations- ‘greenville sc’, ‘greensboro nc’ ‘direct pay fayetteville arkansas’
  2. Names - ‘earl t riggs’
  3. Random terms that offer no query value - ‘are’, ‘cancel’ or ‘chang’
  4. Spelling Errors - ‘ebenefis’, ‘ebeneifits’,’ebenifts’, ‘ebinifits’ and ‘ebnefits’) or ‘careers’s’, ‘dimentia care’ ‘direcci√≥n de oficina de travel’, ’disabili’
* **Once a term has been removed (via blacklisting) how can we prevent it from returning in the future?** 
  1. We understand that the algorithm will continue to identify suggestions from search queries and don’t want to blacklist terms only to have them return over time
* **Can we modify a term on the typeahead suggestion list?**

Examples:

* 1. Change ‘covid’ to ‘Covid-19’
  2. Change ‘v a form 20-0995’ to ‘va form 20-0995’
  3. Change ‘disabiled children’ to ‘disabled children’
* **Can you provide any documentation on the “Did you mean” spelling feature about how it works?**
* **Will a misspelled search query that generates search results** (ex: “[goverment](https://staging.va.gov/search/?query=goverment)” has about 40 hits) produce a “did you mean?” alternate query?