

**Eli Agbayani** <eagbayani@eol.org>

taxon summary method18 messages

Eli Agbayani <eagbayani@eol.org>

Wed, Sep 12, 2018 at 10:58 AM

To: Jen Hammock <jen.hammock@gmail.com>

Hi Jen,

Can you please show me how you create the tree for 'taxon summary' method.

Attached are working files for these two page IDs and predicate:

-- page_id: 46559118 | predicate: [http://purl.obolibrary.org/obo/RO_0002439] (preys on)-- page_id: 7673 | predicate: [http://purl.obolibrary.org/obo/RO_0002470] (eats)I got the ancestry information from this [Dynamic Hierarchy](#).

*Also I noticed that from traits.csv (from Carnivora dataset), that the field 'target_scientific_name' is blank for all rows.

Please tell me if you need other working files, so I can provide it.

Or if you want a specific page ID for our test case.

Thanks

Eli

**Archive.zip**

5K

Jen Hammock <jen.hammock@gmail.com>

Wed, Sep 12, 2018 at 1:39 PM

To: Eli Agbayani <eagbayani@eol.org>

Thanks, Eli! Let's work with 46559118. More records is probably good for testing. For working data, may I please also have the ancestry array (just like you have it for 46559118, for each of the value page_ids?

No scientificName, huh? I guess people can get it from the taxa file, but there is a field for it in the data table, so it ought to be displayed. Thanks, I'll ask Jeremy to fix it.

:)

Jen


[Quoted text hidden]

Eli Agbayani <eagbayani@eol.org>

Wed, Sep 12, 2018 at 3:04 PM

To: Jen Hammock <jen.hammock@gmail.com>

Hi Jen,
Here is the updated 46559118_result.txt
Thanks,
Eli
[Quoted text hidden]

 **46559118_result.txt**
18K

Jen Hammock <jen.hammock@gmail.com>
To: Eli Agbayani <eagbayani@eol.org>

Wed, Sep 12, 2018 at 3:16 PM

Oh, I'm sorry! I thought I was being so precise... I should have asked for the ancestry for all the object_page_ids. My bad! (I don't need any of the other ancestry arrays, actually).

Jen
[Quoted text hidden]

Eli Agbayani <eagbayani@eol.org>
To: Jen Hammock <jen.hammock@gmail.com>

Wed, Sep 12, 2018 at 4:38 PM

Here it is Jen.
Ancestry of object_page_ids for 46559118
Thanks.
[Quoted text hidden]

 **object_page_ids_ancestry.txt**
7K

Jen Hammock <jen.hammock@gmail.com>
To: Eli Agbayani <eagbayani@eol.org>

Wed, Sep 12, 2018 at 5:08 PM

OK, this is easier for me to visualize than to articulate, so here is that example back. The "output" is at the bottom, with columns arranged so you may see patterns by eyeballing it. The idea was to find the common ancestry of all the records and then remove any ancestor nodes that do not add topology to the tree. I think the process goes like this:

For each ancestor, find all recs in which it appears (recs set 1)
If the parent of that ancestor is the same in all the recs in rec set 1, remove the parent

I think that's it, but before we move on, maybe you'd better try it and see how your results compare with mine.

If it works, then we basically have the tree, though I'm not sure if you can use it as is (ordered lists of ancestors) or if it would be better to convert it to parent-child relationships for the next step.

Standing by,

Jen

[Quoted text hidden]

Jen Hammock <jen.hammock@gmail.com>
To: Eli Agbayani <eagbayani@eol.org>

Wed, Sep 12, 2018 at 5:09 PM

sorry, did I not send the file?

[Quoted text hidden]

 **object_page_ids_ancestry.txt**
20K

Eli Agbayani <eagbayani@eol.org>
To: Jen Hammock <jen.hammock@gmail.com>

Thu, Sep 13, 2018 at 4:00 AM

Hi Jen,

Thanks. Here it is. On left is my tip. e.g. 5500.

Notes:

- Some I have more roots than you
- But all your roots, exist in mine

Please check our discrepancy e.g. for tip 5500 below. If I missed something in the procedure. Thanks.

Array

```
(  
  [5500] => Array  
    (  
      [0] => 40054820 - discrepancy  
      [1] => 40054568  
      [2] => 40054103 - discrepancy  
      [3] => 40045861  
      [4] => 2775704  
      [5] => 3014411  
    )  
)
```

```
[46562761] => Array  
(  
  [0] => 3190  
  [1] => 40045861  
  [2] => 2775704  
  [3] => 3014411  
)
```

```
[5419] => Array  
(  
  [0] => 40054568  
  [1] => 40054103  
  [2] => 40045861  
  [3] => 2775704  
  [4] => 3014411  
)
```

```
[46577310] => Array
(
    [0] => 40055115
    [1] => 40054820
    [2] => 40054568
    [3] => 40054103
    [4] => 40045861
    [5] => 2775704
    [6] => 3014411
)
```

```
[46582151] => Array
(
    [0] => 40054103
    [1] => 40045861
    [2] => 2775704
    [3] => 3014411
)
```

```
[695] => Array
(
    [0] => 2775704
    [1] => 3014411
)
```

```
[46562559] => Array
(
    [0] => 2776967
    [1] => 3190
    [2] => 40045861
    [3] => 2775704
    [4] => 3014411
)
```

```
[46563235] => Array
(
    [0] => 40054103
    [1] => 40045861
    [2] => 2775704
    [3] => 3014411
)
```

```
[46562411] => Array
(
    [0] => 2776967
    [1] => 3190
    [2] => 40045861
    [3] => 2775704
    [4] => 3014411
)
```

```
[46578118] => Array
(
  [0] => 40055115
  [1] => 40054820
  [2] => 40054568
  [3] => 40054103
  [4] => 40045861
  [5] => 2775704
  [6] => 3014411
)
```

```
[448838] => Array
(
  [0] => 3014411
)
```

```
[5319] => Array
(
  [0] => 40055115
  [1] => 40054820
  [2] => 40054568
  [3] => 40054103
  [4] => 40045861
  [5] => 2775704
  [6] => 3014411
)
```

)
Thanks,
Eli

[Quoted text hidden]

Jen Hammock <jen.hammock@gmail.com>
To: Eli Agbayani <eagbayani@eol.org>

Thu, Sep 13, 2018 at 12:08 PM

Thanks, Eli! Looks like you were correct in both cases; I just missed them.

:)

Jen

[Quoted text hidden]

Eli Agbayani <eagbayani@eol.org>
To: Jen Hammock <jen.hammock@gmail.com>

Thu, Sep 13, 2018 at 1:45 PM

Great, thanks Jen.

So we now have a tree.

If you can again proceed with the next steps and show me again how you generated it.

That will be great.

Thanks,

Eli

[Quoted text hidden]

Jen Hammock <jen.hammock@gmail.com>

Thu, Sep 13, 2018 at 2:27 PM

To: Eli Agbayani <eagbayani@eol.org>

OK, in doing this example I noticed two steps I had been leaving out. Those are annotated in the attached; all the new action is toward the end.

This one turns out to be resolved at step 2, so if you would like a longer one, may I suggest the same predicate for <https://beta.eol.org/pages/328607>

Let me know if this makes sense as far as it goes, anyway,

Jen

[Quoted text hidden]

Eli Agbayani <eagbayani@eol.org>

Thu, Sep 13, 2018 at 6:09 PM

To: Jen Hammock <jen.hammock@gmail.com>

Hi Jen,


Seems you forgot the attachment.

Also as far as the Carnivora dataset we're using, the page_id = 328607 (see attached) doesn't have trait data for predicate "http://purl.obolibrary.org/obo/RO_0002439" (preys on).

Thanks,

Eli

[Quoted text hidden]

 **328607.txt**
270K

Jen Hammock <jen.hammock@gmail.com>

Thu, Sep 13, 2018 at 7:28 PM

To: Eli Agbayani <eagbayani@eol.org>

Sorry! Here it is.

and you're right about page_id = 328607. The predicate I meant is http://purl.obolibrary.org/obo/RO_0002470

THanks for checking!

Jen

[Quoted text hidden]

 **object_page_ids_ancestry (2).txt**
18K

Eli Agbayani <eagbayani@eol.org>
To: Jen Hammock <jen.hammock@gmail.com>

Thu, Sep 13, 2018 at 7:45 PM

Thanks Jen. Will give you feedback soon.
Thanks,
Eli

[Quoted text hidden]

Eli Agbayani <eagbayani@eol.org>
To: Jen Hammock <jen.hammock@gmail.com>

Fri, Sep 14, 2018 at 5:05 AM

Hi Jen, 2 clarifications please.
1. In your step 1, before the extra step. Why is 40045861 not included.
2. Also when you say **root ancestor** in:
"find set 1: all nodes that are parents of tips, unless *the tip or the parent* is the **root ancestor**, in which case keep the tip in set 1 instead. "

Root is just one and it is = 3014411 .. is that correct?
Or is there a computation to get all roots?

Thanks,
Eli

[Quoted text hidden]

Jen Hammock <jen.hammock@gmail.com>
To: Eli Agbayani <eagbayani@eol.org>

Fri, Sep 14, 2018 at 9:08 AM

Oh cool! I didn't even notice that pattern. So, for step 1: select not all the ancestors, but only the immediate parents, of the tips- so, not the grandparents or great-grandparents, etc. In this case, nearly all of the ancestors are immediate parents of one of the tips- just not that one. Since 40045861 is a grandparent, a great-grandparent, etc., but not an immediate parent, it is excluded.

Roots: yes, after removing any records not in the Dynamic hierarchy, there should always be just one root. (Part of me realized that and part of me forgot.) In this case it is 3014411.

Thanks for these observations. You've helped me realize how much simpler this one can be. Having a single root does change things. I'll update the doc, but basically you can skip right to the last step. After the tree is constructed:

Select all immediate children of the root and label REP.

Label the root PRM

How does that sound? We can still do the page_id=328607 example if you'd like to make sure, but it shouldn't be very different from this one. Construct the tree, find the root and the immediate children.

:)

Jen

[Quoted text hidden]

Eli Agbayani <eagbayani@eol.org>
To: Jen Hammock <jen.hammock@gmail.com>

Fri, Sep 14, 2018 at 2:59 PM

Thanks for the explanation Jen.

Using the new steps:

- Construct the tree,
- find the root
- and the immediate children.

Attached are the results for:


page_id = 46559118; predicate = "http://purl.obolibrary.org/obo/RO_0002439"; //preys on
page_id = 328607; predicate = "http://purl.obolibrary.org/obo/RO_0002470"; //eats


At the bottom are the final results.

Thanks,
Eli

[Quoted text hidden]

2 attachments

 **328607_result.txt**
31K

 **46559118_result.txt**
9K

Jen Hammock <jen.hammock@gmail.com>
To: Eli Agbayani <eagbayani@eol.org>

Fri, Sep 14, 2018 at 3:51 PM

That looks right!

caveat: we seem to be selecting obscure ancestors like "amniota" a lot, so we might tweak this later (say, choose the grandchildren rather than the direct children, or something) but I think I should see a larger sample of results, and consult Katja, before trying to figure that out.

If you're ready to move on, I think the "parents" process, several worksheets to the right, is worth doing next. What you might need first is a way to select taxa for which to make summary records in the first place. I think as a starting point, We should try just doing all taxa with rank=species. I'll bet there will be more ranks worth doing eventually, but Katja and I will need to give that some thought.

Thanks!

Jen

[Quoted text hidden]