# Parsing congress.gov for presidential nominations

Michael Lindsey - Law Library - 2018

CONGRESS.	GOV	Advanced Searches Browse Search Tools Support Sign In * Legislation Congressional Record Committees Members
Nominations 🔶	Examples: PN2	
	Examples. File	
97th-115th Congress (1981-2018)   <u>A</u> Save this Search   Download Resul		BPrint Subscribe Share/Save Sive Feedback
Refined by: Nominations x		
Hide Filters	^	1-100 of 34,726         100 per page           Sort by         Number - Descending           View         Expanded
Limit Your Search Check all	Θ	NOMINATION         1. PN1617 — 115th Congress (2017-2018) — Jon C. Peterson — Marine Corps         Creating intervention in the second section with less C. Peterson
Legislation	[407,498]	One nomination, beginning with Jon C. Peterson, and ending with Jon C. Peterson Date Received from President: 02/08/2018
Members	[2,247]	Committee: Armed Services
Congressional Record	[578,187]	Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services. (All Actions)
Committee Reports	[15,017]	NOMINATION
Nominations	[34,726]	2. PN1616 — 115th Congress (2017-2018) — Marine Corps
Treaty Documents	[768]	2 nominations, beginning with ODIN PINEDA, and ending with JAMES M. ROD
House Communications	[15,438]	Date Received from President: 02/08/2018 Committee: Armed Services
Senate Communications	[147,253]	Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services. (All Actions)
Congress	Θ	NOMINATION
Check all		3. PN1615 — 115th Congress (2017-2018) — Marine Corps
115 (2017-2018)	[1,590]	2 nominations, beginning with MATTHEW C. PAMPUSH, and ending with STEPHEN T. SUTTON Date Received from President: 02/08/2018
114 (2015-2016)	[1,931]	Committee: Armed Services
113 (2013-2014)	[2,237]	Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services. (All Actions)

Limit Your Search  I PN161Z - 115th Congress (2017-2018) — Jon C. Peterson — Marine Corps Doe nomination, beginning with Jon C. Peterson , and ending with Jon C. Peterson )  Pate Received from President: 02/08/2018 Committee: Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services Latest Action: 02/08/20			· · · · · · · · · · · · · · · · · · ·	
Check all Check	Limit Your Search	Ξ	NOMINATION	
One nomination, beginning with Jon C. Peterson, and ending with Jon C. Peterson         Legislation       [407,498]         Members       [2,247]         Compressional Record       [578,197]         Committee Reports       [15,017]         Nominations       [94,726]         Nominations       [94,726]         Nominations       [94,726]         Nominations       [94,726]         Pressional Record       [15,017]         Nominations       [94,726]         Pressional Record       [15,017]         Nominations       [94,726]         Pressional Record       [15,017]         Nominations       [16,16]         House Communications       [16,438]         Committee: Armed Services       Latest Action: 02/08/2018         Communications       [147,253]         Latest Action: 02/08/2018 - Received from President: 02/08/2018         Communications       [147,253]         Latest Action: 02/08/2018 - Received from President: 02/08/2018         Console       Debugger         Inspector       Console         Debugger       [15,152,7721]         Vertic Lass="result-teading">	Chock all		1. PN1617 — 115th Congress (2017-2018) — Jon C. Peterson — Marine Corp	S
Members       [2,247]         Members       [2,247]         Congressional Record       [578,157]         Committee And Services       Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services         Committee Reports       [15,017]         Nominations       [34,726]         Treaty Documents       [768]         Prominations, beginning with ODIN PINEDA, and ending with JAMES M. ROD         Date Received from President: 02/08/2018         Communications       [147,253]         Inspector       Console         Debugger {} Style Editor       © Performance         Inspector       Console         Inspector       Conspress, qov/nomination/115th-congress/16172r=			One nomination, beginning with Jon C. Peterson, and ending with Jon C. Peterson	
Image: Nominations       (2,247)         Congressional Record       (527,137)         Committee Reports       (15,017)         Nominations       (34,726)         Image: Nominations       (34,726)         Treaty Documents       (768)         House Communications       (15,438)         Senate Communications       (147,253)         Inspector       Console         Debugger       {} Style Editor         @ Performance       {} Memory         Inspector       Console         Debugger       {} Style Editor         @ relations-results-Lists expanded-view' start='1'>         <	Legislation	[407,498]	Date Received from President: 02/08/2018	
Congressional Record       [15,017]         Nominations       [14,728]         Treaty Documents       [16,438]         House Communications       [15,438]         Senate Communications       [147,253]         Inspector       © Console         Debugger       () Style Editor         @ Performance       [] Memory         > <ol> <li></li> <li></li></ol>	Members	[2,247]	Committee: Armed Services	
Wominations       [34,726]         Nominations       [34,726]         Treaty Documents       [768]         House Communications       [15,438]         Senate Communications       [147,253]         Inspector       Console         Debugger       {} Style Editor         @ Nominations       [147,253]         Inspector       Console         Debugger       {} Style Editor         @ Performance       @ Memory         @ Search HTML       @ Rules         * <<01 class="result-heading">         * <<01 class="result-heading">         * <<01 class="result-heading">         * <<01 class="result-tiem">         * <01 class="result-heading">         * <01 class="result-heading">         * <02 class="result-heading">         * <02 class="result-heading">         * <02 class="result-tiem">         * <02 class="result-ti		[578,187]	Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed	d Services.
<pre>Introduct Construction Construction Construction Construction Construction Construction Construction Constructions Construc</pre>	Committee Reports	[ <mark>15,017]</mark>	NOMINATION	
Treaty Documents       [768]         Treaty Documents       [768]         House Communications       [15,438]         Senate Communications       [147,253]         Inspector       Console       Debugger {} Style Editor       @ Performance       Memory       > Network       @ Storage         Inspector       Console       Debugger {} Style Editor       @ Performance       @ Memory       > Network       @ Storage         Inspector       Console       Debugger {} Style Editor       @ Performance       @ Memory       > Network       @ Storage         Inspector       Console       Debugger {} bucks!">       Rules         Inspector       Console       Debugger {} bucks!">       Rules         Inspector       Console       Debugger {} bucks!">       Rules         Inspector       Console       Style Editor       @ Performance       Memory       > Network       @ Storage         Inspector       Console       Debugger {} bucks!">       Rules         Inspector       Console       Storage       @ Flite         Issaic-search-results-lists       expanded-view" start="1">         Issaic-search-results-lists       expandelseine       .basic-search-results         Isth       Congress       (2	Nominations	[34,726]	2 PN1616 - 115th Congress (2017-2018) - Marine Corps	
House Communications       [15,438]         Date Received from President: 02/08/2018         Committee: Armed Services         Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services         Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services         Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services         Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services         Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services         V <ol class="basic-search-results-lists expanded-view" start="1">         V<ol class="result-lists://www.congress.gov/nomination/115th-congress/1617?r=1">         V<ol class="result-heading">         &lt;</ol></ol></ol></ol></ol></ol></ol>				
House Communications       [15,438]         Senate Communications       [147,253]         Inspector       Console       Debugger       {} Style Editor       @ Performance       @ Memory       > Network       @ Storage         Inspector       Console       Debugger       {} Style Editor       @ Performance       @ Memory       > Network       @ Storage         Inspector       Console       Debugger       {} Style Editor       @ Performance       @ Memory       > Network       @ Storage         Inspector       Console       Debugger       {} Style Editor       @ Performance       @ Memory       > Network       @ Storage         Inspector       Console       Debugger       {} Style Editor       @ Performance       @ Memory       > Network       @ Storage         Inspector       Console       Debugger       {} Style Editor       @ Performance       @ Memory       > Network       @ Storage         Inspector       Class="expanded-view" start="1">       @ Search HTML       @ Performance	Treaty Documents	[768]		
Senate Communications       [147,253]         Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed Services.         Inspector       E Console         Debugger       {} Style Editor         @ Search HTML       Rules	House Communications	[15,438]		
<pre>Inspector Console Debugger {} Style Editor @ Performance @ Memory = Network @ Storage</pre>		14 47 0503		
Q. Search HTML       ✓         ▼ <ol class="basic-search-results-lists expanded-view" start="1">         ▼<li class="expanded" style="display: block;">         ▼<li class="expanded" style="display: block;">         ▼<li>          ↓       <!--</th--><th></th><th>[,1200]</th><th>Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed</th><th>1 Services</th></li></li></li></ol>		[,1200]	Latest Action: 02/08/2018 - Received in the Senate and referred to the Committee on Armed	1 Services
<pre>v<li>class="expanded" style="display: block;"&gt;</li></pre>				Rules
<pre>&gt; <div> = </div> =  =  =  =  =  =  =  =  =   =  <td></td><td></td><td></td><td></td></pre>				
<pre>i. did i. did v <span class="result-heading"></span></pre>				V Filte
<pre>&gt;&gt; <span class="result-heading"> }</span></pre>	and the provide the second	ALCONOM AND A DAMAGE AND A	otay: block;">	1
<pre><a href="https://www.congress.gov/nomination/115th-congress/1617?r=1">PN1617</a>  <pre>.basic- astrong&gt;Jon C. Peterson - Marine Corps basic- astrong&gt;Dates="result-item"&gt; basic- astrong&gt;Date Received from President: basic- ba</pre></pre>		ALCONOM AND A DAMAGE AND A	Juay: DLOCK;">	element
<pre><strong>Jon C. Peterson - Marine Corps</strong> ::after  &gt; <span class="result-item">=</span> inherited  inherited  iou class="result-item"&gt;= iou class="result-item"&gt;&gt;= iou class="result-item"&gt;&gt;</pre>			Juay: DLOCK;">	element
<pre>::after  &gt; <span class="result-item"></span></pre> Inheriter Inheriter Inheriter Inheriter Inheriter Inheriter Inheriter Inheriter Inheriter Inheriter Inheriter Inheriter Inheriter Inheriter Inheriter	<pre>w<span class="r&lt;br&gt;&lt;a href=" http:="" second="" second<="" td="" www.second.com=""><td>esult-heading"&gt; ps://www.congress</td><td>s.gov/nomination/115th-congress/1617?r=1"&gt;PN1617</td><td>element di } .basic-</td></span></pre>	esult-heading"> ps://www.congress	s.gov/nomination/115th-congress/1617?r=1">PN1617	element di } .basic-
<pre>   </pre>	<pre>▼<span <u="" class="r&lt;br&gt;&lt;a href=">http - 115th Congr</span></pre>	esult-heading"> <u>ps://www.congress</u> ress (2017-2018)	s.gov/nomination/115th-congress/1617?r=1">PN1617	element di } .basic- search-
<pre>     <span class="result-item"></span></pre>	<pre>▼<span class="r&lt;br&gt;&lt;a href=" http:<br="">- 115th Cong <strong>Jon (</strong></span></pre>	esult-heading"> <u>ps://www.congress</u> ress (2017-2018)	s.gov/nomination/115th-congress/1617?r=1">PN1617	element di } .basic- search- results
<pre></pre>	<pre>▼<span class="r&lt;br&gt;&lt;a href=" http<br="">- 115th Cong <strong>Jon ( ::after</strong></span></pre>	esult-heading"> <u>ps://www.congress</u> ress (2017-2018)	s.gov/nomination/115th-congress/1617?r=1">PN1617	element di } .basic- search- results ma
<pre><strong>Date Received from President:</strong>     search-     02/08/2018       li     <span class="result-item">=</span>     }      /li&gt; </pre>	<pre>span class="r <a href="http&lt;br&gt;- 115th Cong&lt;br&gt;&lt;strong&gt;Jon (&lt;br&gt;::after&lt;br&gt;&lt;/span&gt;&lt;/pre&gt;&lt;/td&gt;&lt;td&gt;esult-heading"> <u>ps://www.congress</u> ress (2017–2018) C. Peterson — Mar</a></pre>	s.qov/nomination/115th-congress/1617?r=1">PN1617 - rine Corps	element di } .basic- search- results ma }	
<pre>\tits  &gt; \tits = } </pre>	<pre><span class="r&lt;br&gt;&lt;a href=" httg<br="">- 115th Cong <strong>Jon ( ::after </strong></span> &gt;<span class="r&lt;/pre&gt;&lt;/td&gt;&lt;td&gt;esult-heading"> <u>ps://www.congress</u> ress (2017–2018) C. Peterson — Mar esult-item"&gt;—<!--</td--><td>s.qov/nomination/115th-congress/1617?r=1"&gt;PN1617 - rine Corps</td><td>element di } .basic- search- results ma } Inherite</td></span></pre>	s.qov/nomination/115th-congress/1617?r=1">PN1617 - rine Corps	element di } .basic- search- results ma } Inherite	
<pre>&gt; <span class="result-item"></span></pre> > } > <span class="result-item">&gt; //// </span>	<pre><span class="r&lt;br&gt;&lt;a href=" httg<br="">- 115th Cong <strong>Jon ( ::after </strong></span> &gt;<span class="r&lt;br&gt;&lt;span class=" r<br=""><strong>Date</strong></span></pre>	esult-heading"> <u>ps://www.congress</u> ress (2017-2018) C. Peterson — Mar esult-item">(=) </td <td>s.qov/nomination/115th-congress/1617?r=1"&gt;PN1617 - rine Corps /span&gt;</td> <td>element di } .basic- search- results ma } Inherite .basic-</td>	s.qov/nomination/115th-congress/1617?r=1">PN1617 - rine Corps /span>	element di } .basic- search- results ma } Inherite .basic-
<pre>&gt; <span class="result-item">=&gt; </span>  </pre>	<pre>\$</pre>	esult-heading"> <u>ps://www.congress</u> ress (2017-2018) C. Peterson — Mar esult-item">(=) </td <td>s.qov/nomination/115th-congress/1617?r=1"&gt;PN1617 - rine Corps /span&gt;</td> <td>element di } .basic- search- results ma } Inherited .basic- search-</td>	s.qov/nomination/115th-congress/1617?r=1">PN1617 - rine Corps /span>	element di } .basic- search- results ma } Inherited .basic- search-
	<pre><span class="r&lt;br&gt;&lt;a href=" httg<br="">- 115th Cong <strong>Jon ( ::after </strong></span> &gt; <span class="r&lt;br&gt;&lt;strong&gt;Date&lt;br&gt;02/08/2018&lt;br&gt;&lt;/span&gt;&lt;/pre&gt;&lt;/td&gt;&lt;td&gt;esult-heading"> <u>ps://www.congress</u> ress (2017-2018) C. Peterson – Mar esult-item"&gt; esult-item"&gt; esult-item"&gt; Received from Pr</span></pre>	<pre>s.gov/nomination/115th-congress/1617?r=1"&gt;PN1617 rine Corps /span&gt; resident:</pre>	element di } .basic- search- results ma } Inherited .basic- search- results	
	<pre><span class="r&lt;br&gt;&lt;a href=" httg<br="">- 115th Cong <strong>Jon ( ::after </strong></span> &gt; <span class="r&lt;br&gt;&lt;strong&gt;Date&lt;br&gt;02/08/2018&lt;br&gt;&lt;/span&gt;&lt;br&gt;&gt; &lt;span class=" pre="" r<=""></span></pre>	esult-heading"> ps://www.congress ress (2017-2018) C. Peterson – Mar esult-item"> () <br esult-item"> () Received from Pr result-item"> ()	<pre>s.gov/nomination/115th-congress/1617?r=1"&gt;PN1617  rine Corps /span&gt; resident:</pre>	element di } .basic- search- results ma } Inherited .basic- search- results
	<pre>v <span class="r&lt;br&gt;&lt;a href=" httg<br="">- 115th Cong <strong>Jon ( ::after </strong></span> &gt; <span class="r&lt;br&gt;v &lt;span class=" r<br="">v <span class="r&lt;br&gt;v &lt;span class=" r<br="">v <span class="r&lt;br&gt;v &lt;span class=" r<br="">v <span class="r&lt;br&gt;v &lt;span class=" pre="" r<=""></span></span></span></span></pre>	esult-heading"> ps://www.congress ress (2017-2018) C. Peterson – Mar esult-item"> () <br esult-item"> () Received from Pr result-item"> ()	<pre>s.gov/nomination/115th-congress/1617?r=1"&gt;PN1617  rine Corps /span&gt; resident:</pre>	<pre>element    di    di    .basic- search- results    ma } Inherited .basic- search- results</pre>

## step one:

get links to all the nominations

• base url:

https://www.congress.gov/search?q={%22source%22:%22nominations%22,%22nomination-t
ype%22:%22Civilian%22}&pageSize=250

- set \$pageNumber = 1
- tack **\$pageNumber** onto the end of the base url like so: ... **&page=\$pageNumber** and grab it over HTTP.
- This returns a list of nomination in batches of 250
- Convert that HTML to XML and break the list up into individual nominations
   \$1ist = \$xpath->query("//ol[contains(@class,'basic-search-results-lists')]
   /li[contains(@class,'compact')]/span[contains(@class,'result-heading')]/a");
- for each item in that list, harvest the link (href) and add a new nomination to the database
- if there's a link to a next page of nominations, increment **\$pageNumber** and do it all again

Now we have a database of 34K+ nominations, but so far, just the links to their pages.

### All Information for PN1616 — Marine Corps

115th Congress (2017-2018) | Get alerts

### « Back to this nomination

Nominees	Latest Action
There are 2 nominations, beginning with ODIN PINEDA, and ending with JAMES M. ROD	02/08/2018 - Received in the Senate and referred to the Committee on Armed Services.
Position	Date Received from President
To be Major	02/08/2018
Organization	Committee
Marine Corps	Senate Armed Services

### Jump to: Actions | Nominees

### Actions (1)

Date	Senate Actions
02/08/2018	Received in the Senate and referred to the Committee on Armed Services.

### Nominees (2)

THE FOLLOWING NAMED LIMITED DUTY OFFICERS FOR APPOINTMENT TO THE GRADE INDICATED IN THE UNITED STATES MARINE CORPS UNDER TITLE 10, U.S.C., SECTION 624: To be Major

### Nominee

ODIN PINEDA JAMES M. ROD

### step two:

### get the raw HTML for each nomination and store it

- grab references to all nominations in the database where we don't have the raw HTML (that will be all of them at first)
- in a loop, follow the link and store the HTML there for that particular nomination Sometimes this fails for some network reason or other
- re-run this process a few times until there are no more to get

Now we have the raw HTML for every nomination.

mysql> desc nonMilitaryNominations;

idint(11)NOPRINULLauto_incrementnomineetextYESNULLcongresstinyint(3) unsignedYESNULLpnsmallint(5) unsignedYESNULLparttinyint(3) unsignedYESNULLl parttinyint(3) unsignedYESNULLorganizationtextYESNULLorganizationvarchar(255)YESNULLlinkvarchar(255)YESNULL	+   Field	Туре	Null	+   Key	Default	+ Extra
pagemediumtextYESNULLcommitteevarchar(255)YESNULLlastActionDatedateYESNULLlastActionvarchar(255)YESNULLfschar(1)YESN	<pre>nominee congress pn part description position organization dateRecieved link page committee lastActionDate lastAction</pre>	<pre>text tinyint(3) unsigned smallint(5) unsigned tinyint(3) unsigned text text varchar(255) date varchar(255) mediumtext varchar(255) date varchar(255)</pre>	YES YES YES YES YES YES YES YES YES YES	PRI	NULL NULL NULL NULL NULL NULL NULL NULL	auto_increment

-----

\_ \_ \_ \_ \_ \_ \_ \_ \_

15 rows in set (0.01 sec)

### step three:

parse each nomination and store the data we're interested in.

- fetch from the database a batch of nominations where we have the HTML, but haven't yet populated the metadata
- The HTML should either be for a single-nominee, or for multiple nominees under the same PN#. The HTML for both types conforms to a particular form, and we can use regular expressions to tell them apart.
- grab the nominee name(s) from the HTML as well as the other details and store them in the database



- iterate through every nomination in the database and output it in a form that Excel or Google Sheets can understand
- deliver to Anne

# gotchas

### of course there are

- the data format might change slightly at any given time. So build in integrity checks at every point and have the script konk out with a meaningful error message that you can follow up on
- Be prepared to re-write code that made more sense to you a long time ago than it does now
- data might be coded in ways you don't expect. Anne is interested in certain non-military nominations that happen to be included in that dataset.

## last bits

- important to have a clear role in the partnership. Anne has the domain expertise. Concentrate on getting her the data and make as few assumptions as possible.
- ProPublica is also interested in this data and has implemented an API for it, but it currently can't handle nominations of multiple people or committees for those nominations in any useful way. It's marked as an enhancement for a future version of their API.
- JSON or some other structured data format is the natural way for machines to work with data provided by other machines. One day!