Kavanaugh from 09-13-18 to 10-11-18

Government shutdown from 12-08-18 – 02-08-19

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Links | Issues | Solved? | Programmed | Downloaded |
| NY Times | Show more button | Selenium using click element | Kavanaugh  Shut down | BOTH |
| NPR | Can’t get to the links because they are part of script  Page numbers show 20 items at a time, but can loop through the page number  No date range selection | Click element and scrolling to make the page realize you were there.  Sorted dates by using the URLs. | Kavanaugh  Shut down | BOTH |
| Slate | Links might be part of java script  10 results per page; 4,680 total results; no “Next” button for navigating to next results page  Can sort by date, but no range selection | This switched to a google search thing, so was able to click on the numbers at the bottom to get the next pages.  Also used the tag pages because the google search versions were limited to 10 pages only. | Kavanaugh  Shut down | BOTH |
| Huffington Post | Can grab URLS fine, but will need to get Selenium to cycle through pages; There’s a “Next” button, thankfully.  For loop is not grabbing actual text at this point, but it gets the other elements of the data frame.  No sort options; 937 results with 10 per page. | Was able to loop through selenium and hit next key  Pages seem to be blocking grabbing text. Used some sneaks to open the pages and pull the right information. | Kavanaugh  Shut down | BOTH |
| Politico | Links come through fine, but we get this error when running for loop: “Error in open.connection(x, "rb") : HTTP error 401.” The unauthorized error? Surely, we don’t need a login…  There’s a “Next page” button, 2,070 results, and 20 results per page | DOES LET YOU PICK DATES WOO.  Deleted the subscriber information ones. | Kavanaugh  Shut down | BOTH |
| Fox News | Can choose date range in URL; There’s a clickable next page arrow  Article URLs not getting listed. CSS issue?  1,651 results w/ 10 per page. | Figured out how to pull the article lists, cleaned out videos and random search tag pages. | Kavanaugh  Shut down | BOTH |
| Breitbart | Can sort by date, but it brings up weird stuff; can cycle by page in URL  Not getting article URLs; No “next” button, just page numbers.  26,100 results w/ 8 results per page. | Going to use the tag page, the google embedded searches are craptastic  https://www.breitbart.com/tag/brett-kavanaugh/ | Kavanaugh  Shut down | BOTH |
| Rush Limbaugh | Possible java script nonsense preventing link scraping  Little next page arrow for getting to other results  351 results w/ 20 per page. | Clicked on the next icon until done | Kavanaugh  Shut down | BOTH |
| The Blaze | “Error in grep(urlslist9) : argument "x" is missing, with no default”  There’s a “Load More Stories” button  Does not say number of search results. Loads 13 at first, and adds 5 more each time button is clicked | Worked great for the Kavanaugh, but very few articles found for the shutdown (the most recent one) | Kavanaugh  Shut down | BOTH |
| Sean Hannity | Javascript prevents link scraping or url changes.  12 results per page and 8 pages (196 total) | Solution is to the “trending” page, which is like using the tag pages on other sites – also tried to cross reference with google. | Kavanaugh  Shut down | BOTH |