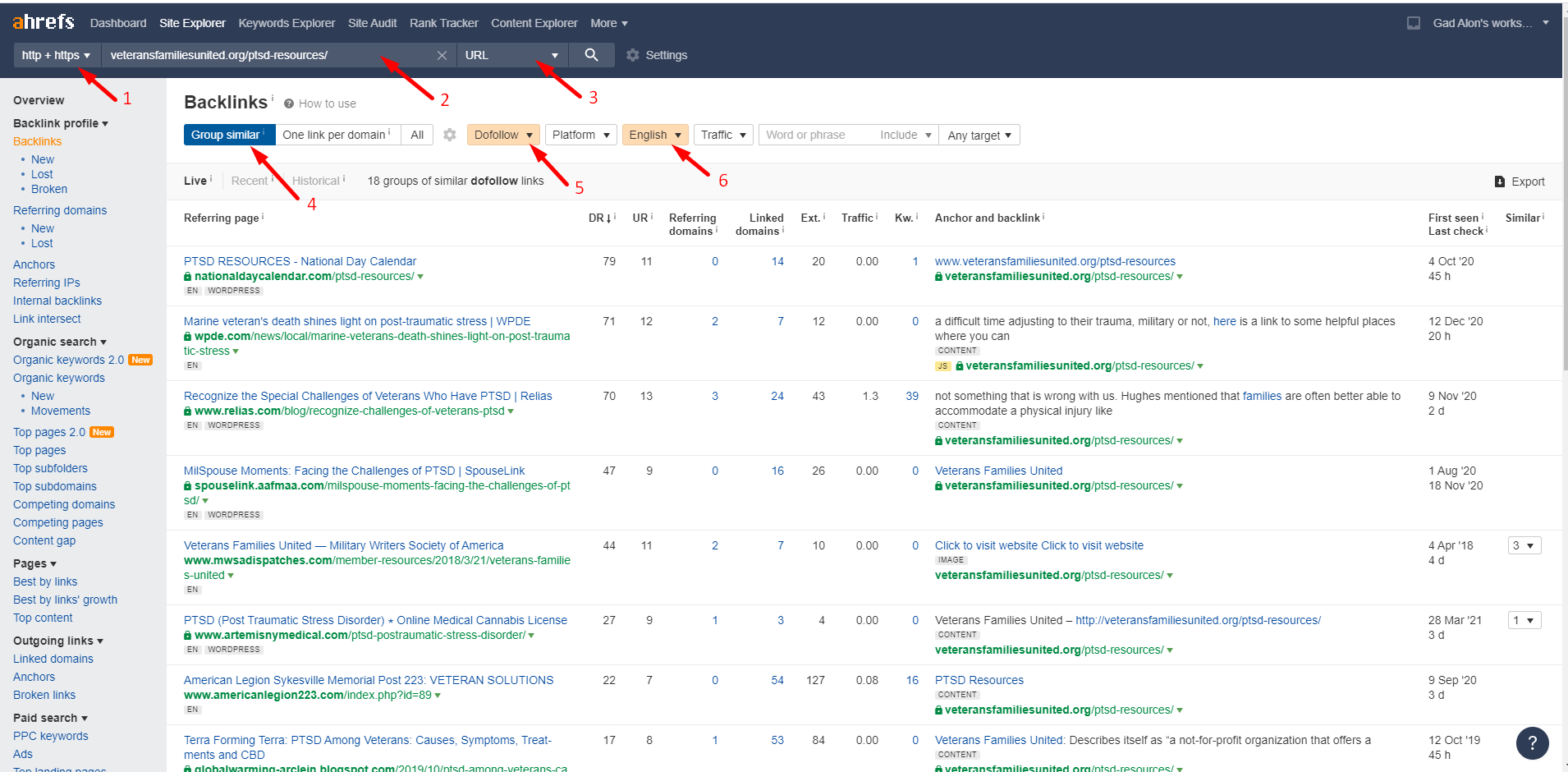
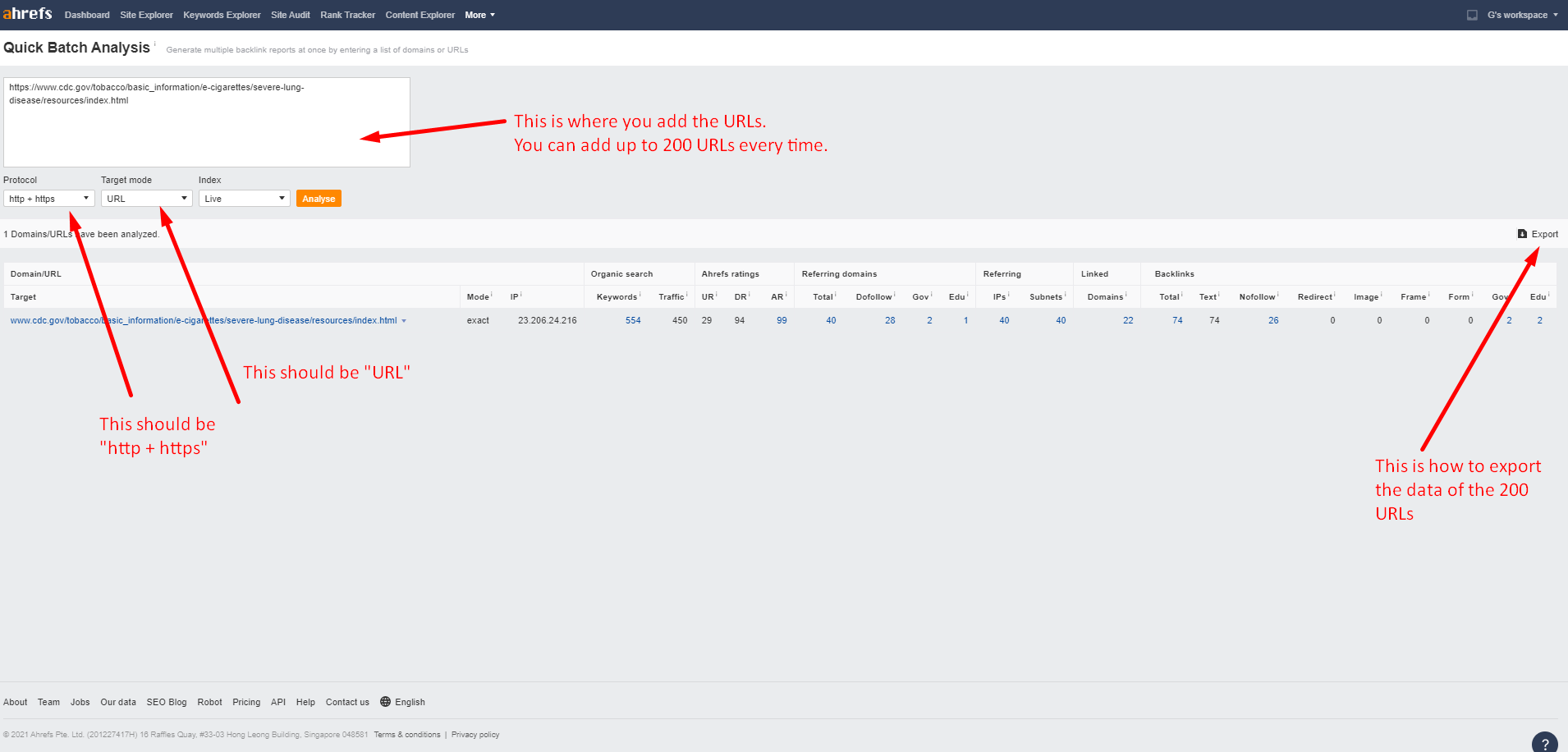
* **Expected output:** 
  + 3 scripts that I will be able to run from my local computer (details below)
    - The first script will take input from **input.xls** and will generate the output file **output-1.xls**
    - The second script will take input from **input.xls** and will generate the output file **output-2.xls**
    - **Then merge output-1.xls** and **output-2.xls** to one file >> **output-3.xls**
    - The third script will take **output-3.xls** and will generate the output file **output-4.xls**
  + Clear instructions on how to run the script on my local computer
* **Input:**
  + An Excel spreadsheet with a list of URLs (see an example file here: **input.xls**)
  + I will provide the credentials (log-in & password) to the **Ahrefs** account
* **Required skills:**
  + Website scraping
  + Familiarity with [Ahrefs](https://ahrefs.com/) interface and functionality.
    - I currently have the [“Lite” subscription](https://ahrefs.com/plans-pricing) to Ahrefs and would hope that this license would suffice for this script.
* **Steps Instructions**
  + **Script #1 –**
  + The script will accept input.xls (as in the example format), and then run **all the URLs** from input.xls in Ahrefs, **one by one**, in the following way. Here is an example for the url: *veteransfamiliesunited.org/ptsd-resources/*
* 
* The URL for the page in the screenshot above is: [*https://ahrefs.com/site-explorer/backlinks/v7/external-similar-links/exact/live/en/all/dofollow/1/domain\_rank\_desc?target=veteransfamiliesunited.org%2Fptsd-resources%2F*](https://ahrefs.com/site-explorer/backlinks/v7/external-similar-links/exact/live/en/all/dofollow/1/domain_rank_desc?target=veteransfamiliesunited.org%2Fptsd-resources%2F)
* Note the IMPORTANT following information in the filters on that page:
  + 0 – you can get to this page by selecting “Backlinks” on the left-hand side menu, after you type in the URL.
  + 1 – this should be “http + https”
  + 2 - this should be the URL without http:// or https://
  + 3 - select “URL” or “Exact URL”
  + 4 – select “Group Similar”
  + 5 – select “DoFollow”
  + 6 – select “English”
* This list should be exported into a spreadsheet using the button/link on the right “Export”. The example output from this export is in the file “export\_ahrefs\_example.csv”.
* This is an export from only ONE of the URLs. You will have an export for each of the URLS from input.xls. There may be an Ahrefs API that can enable to do that faster, but I am unsure if that is available via the current “Lite” subscription.
* merge all the csv files into a single master Excel file >> **output-1.xls**
* **Script #2 –**
* Take all the URLs that are in **input.xls** and use <https://ahrefs.com/batch-analysis> , where you add them URLs (200 URLs at a time) and export the results.
* Gather all the results in **output-2.xls**
* 
* NOW – merge output-1.xls and output-2.xls into a single file **output-3.xls**
* Please pay attention to the column headers when you merge the two spreadsheet. The columns should be matching, but some will be missing and that’s OK.
* **Script #3 –**
* NOW – I want to be able to run a second script on **output-3.xls**. I want to be able to filter the single master file based on the following filters that I want to be able to edit (this means that I can also choose not to use a certain filter):
  + remove duplicate URLs from the column named “Referring Page URL”
  + remove URLs from “Referring Page URL” that contain any keyworks from a list of blacklist keywords.
  + Remove URLs that have more than 3 hyphens. Example: [www.xyz.com/qwe-qwe-qwe-qwe-qwe](http://www.xyz.com/qwe-qwe-qwe-qwe-qwe)
  + Remove URLs that have more than 20 characters after a query string. Query sting starts with “?”. Example: [www.xyz.com/?category=sdfsdfsdf&asdasdas&asdasdasd&asdasdasd](http://www.xyz.com/?category=sdfsdfsdf&asdasdas&asdasdasd&asdasdasd)
* Generate **output-4.xls**