Manual to conduct a media visibility analysis with LexisNexis/SPSS (count keywords in articles)

Rens Vliegenthart (May 16, 2012)

Beta version – please send comments to r.vliegenthart@uva.nl

This manual is meant as a guide on how to conduct a media visibility analysis. It shows how to collect data from LexisNexis, import them into SPSS and make them useable for analysis. The output consists for example of graphs that show the monthly attention for a certain actor in various newspapers. The manual assumes a connection with the digital services of the University of Amsterdam, but it should work largely the same for other universities that provide access to LexisNexis.

The first step you have to take is collecting newspaper articles that are relevant for your research. The first question you have to answer is of course where you're interested in: do you want to investigate the visibility of certain politicians or political issues? And in what media outlets? Until a couple of years ago, conducting a content analysis of newspapers first and foremost meant going into the archives, reading through paper versions of the newspaper and select relevant articles manually. Nowadays, digital databases have made life a lot easier and it is relatively easy to collect large amounts of articles. This works using search terms (combinations of words that need to appear (or not) in the article) and identifying the period and outlets, after which the database selects the articles that match the criteria. The most used database is probably LexisNexis. The academic, Dutch version includes a lot of international, national and local newspapers. For some, only recent years are available, but for some of the media, all articles until back into the nineties or even eighties are included in the database.

(1) When connected to the UvA-web you can get access to LexisNexis via http://academic.lexisnexis.nl/uva/. Change the language in the upper right corner to English' (arrow 1 in picture below. NOTE this is very important – if you use another language, you will run into problems at a later stage) and choose in the upper left corner 'Power Search' (arrow 2 in picture below). The following screen will appear.
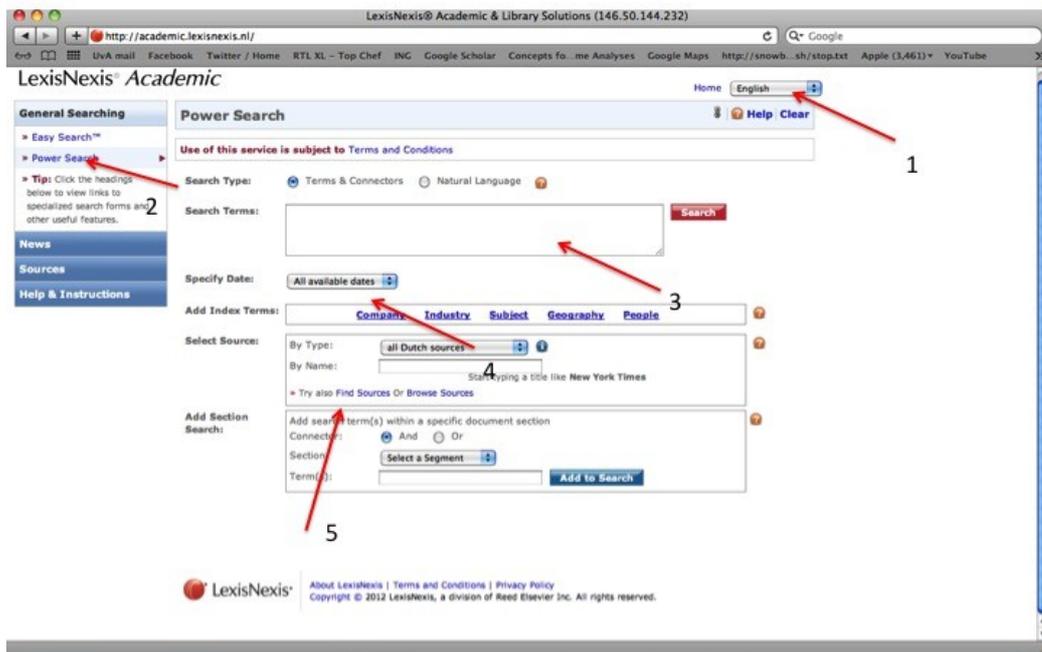
Several choices have to be made.
- Which search terms to use.
- Which period to include.
- Which sources to include.

(2) Establishing the appropriate search terms is probably the most important step in the research. On the one hand, you want to try to include as many articles that are relevant as possible. On the other hand, you want to insure that your selection does not include a lot of articles that turn out not to be relevant. Getting the search terms right is mainly a 'trial-and-error' process, where you try several options, look at the results and adjust your search terms accordingly.

The most obvious search term is one simple word. Typing in 'Bush' will yield all articles in which the word 'Bush' occurs once or more times. A search string can also consist of different words that are connected by 'connectors'. Three often used connectors are 'OR', 'AND', 'AND NOT' and 'w/10'.

- The search term 'Bush OR Gore' results in all articles that include either 'Bush' or 'Gore' or both.
- The search term 'Bush AND Gore' results in articles that include both words 'Bush' and 'Gore'.
- The search term 'Bush AND NOT Gore' results in articles where 'Bush' occurs without 'Gore' occurring in the article.
- The search term 'Bush w/10 Gore' results in articles where 'Bush' and 'Gore' occur within a ten-word distance. 'w/5' gives similar results, but then within 5 words distance.

Furthermore, the are several signs that might be useful:

- Use the exclamation mark to replace an infinite number of letters following a word root. You may use only one exclamation mark in a word at the end of a word root.
- Use the asterisk to replace a character in a word. You can use more than one asterisk in a word. You can use it anywhere, except as the first letter.
- With brackets you can partial out parts of the search terms that belong together. You can make it as difficult as you want. For example (islam! OR muslim!) AND (threat! OR terrorism) results in articles in which 'islam' (maar but also for example 'islamists') occurs together with 'threat' (but also with threatening) or with 'terrorism'. The same goes for 'muslim' (but again also for muslims).

The final search term can be inserted in the search field (arrow 3 in picture above).

The second choice that needs to be made is which period to include. This of course depends on your research question. In many cases, it is interesting to consider a somewhat longer period in time, by for example including some time before 9/11 and after 9/11. You can include the dates in the date fields (arrow 4 in picture above).

The third decision to be made is what sources to select. You can choose a standard set (e.g. 'all Dutch newspapers'), but you can also select separate sources by clicking 'find sources' (arrow 5 in picture above) and searching and adding the ones you want to include. Below is a screenshot of a form that is filled out completely.
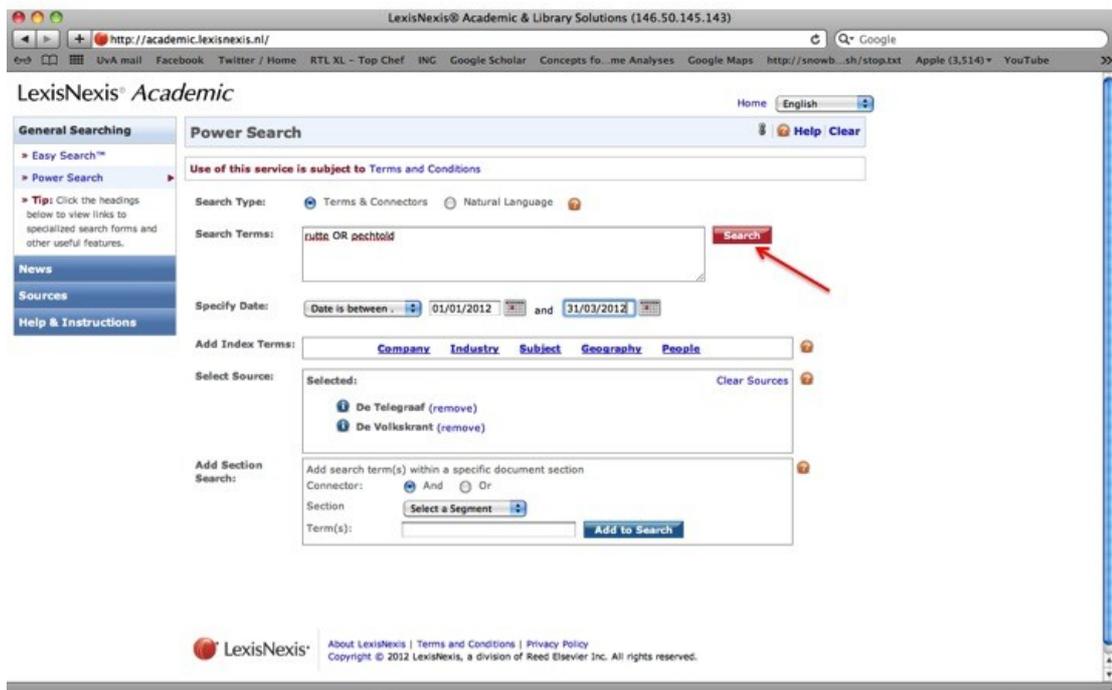

Hereafter, you press the search (arrow in the picture above) button, which gives the following results:
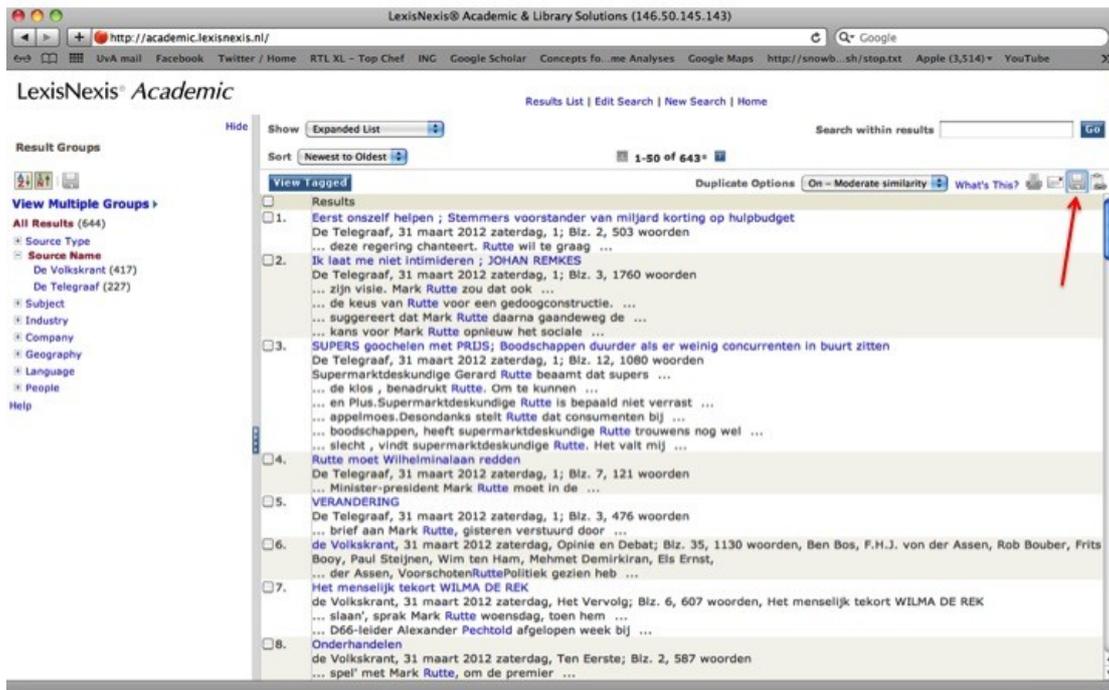
In some cases, LexisNexis indicate that the search results in more than 3000 articles and that those can not be displayed. That's no problem and you can circumvent it by splitting up the time period in for example years and repeat the search action several times. You can now read the articles by clicking on them. By randomly looking at articles you can try to establish whether there are many irrelevant articles included. If that's the case, try to adjust your search term to reduce their number. If you're satisfied with the results, check 50 randomly selected articles and note the number of relevant articles out of those 50.

(3) As a third step, you save the list of articles, by clicking on the download button in the right upper corner (see arrow in picture above). Select the HTML-format and document view 'list'. Keep the rest unchanged and save the file (this might take some time). If you've split up the time period, repeat the procedure and save the results in separate documents.

(4) For SPSS to be able to read the file, it needs to be saved in an excel (.xls/.xlsx) format. This is what you do: (1) Open Excel; (2) Click 'File' and then 'Open'; (3) Select as file type 'All Files'; (4) Select the file you saved from LexisNexis; (5) Open the file and then 'Save', save it as an Excel-file; (6) Repeat those steps for all lists you downloaded; (7) Close Excel!

(5) Now you can proceed to SPSS. Open the syntax that is provided along with this manual. The screen will look as follows:
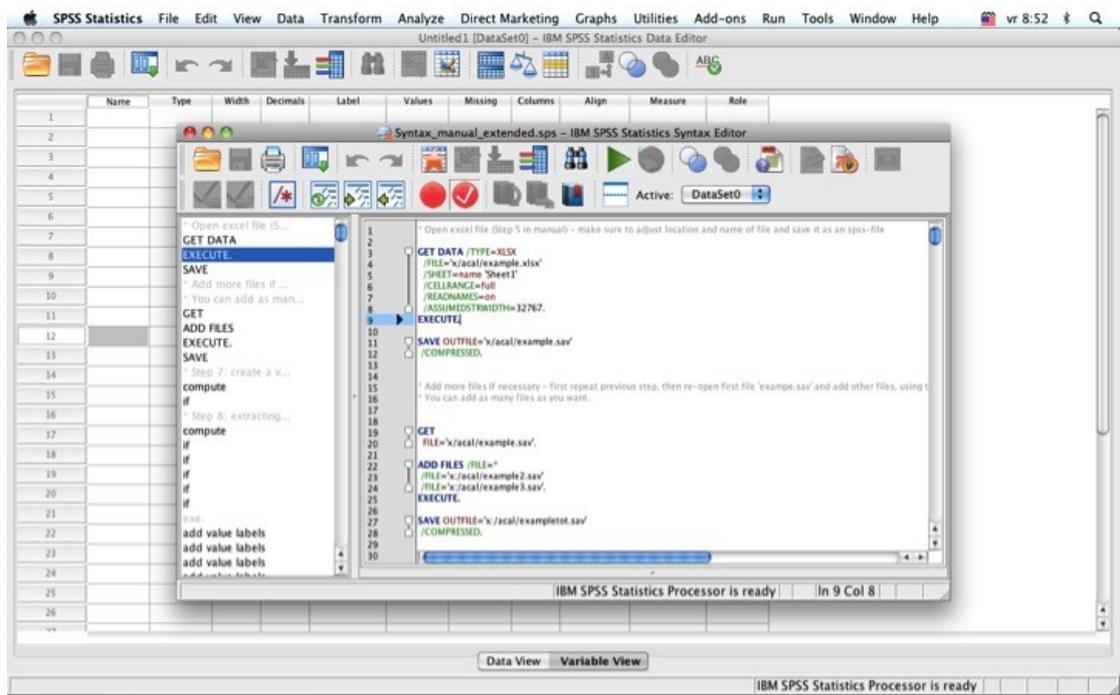
The first part of the syntax opens the excel-file(s) in SPSS. Select it until the next step and press CTRL+R. Make sure that you adjust the name and location of the file. The screen looks as follows.
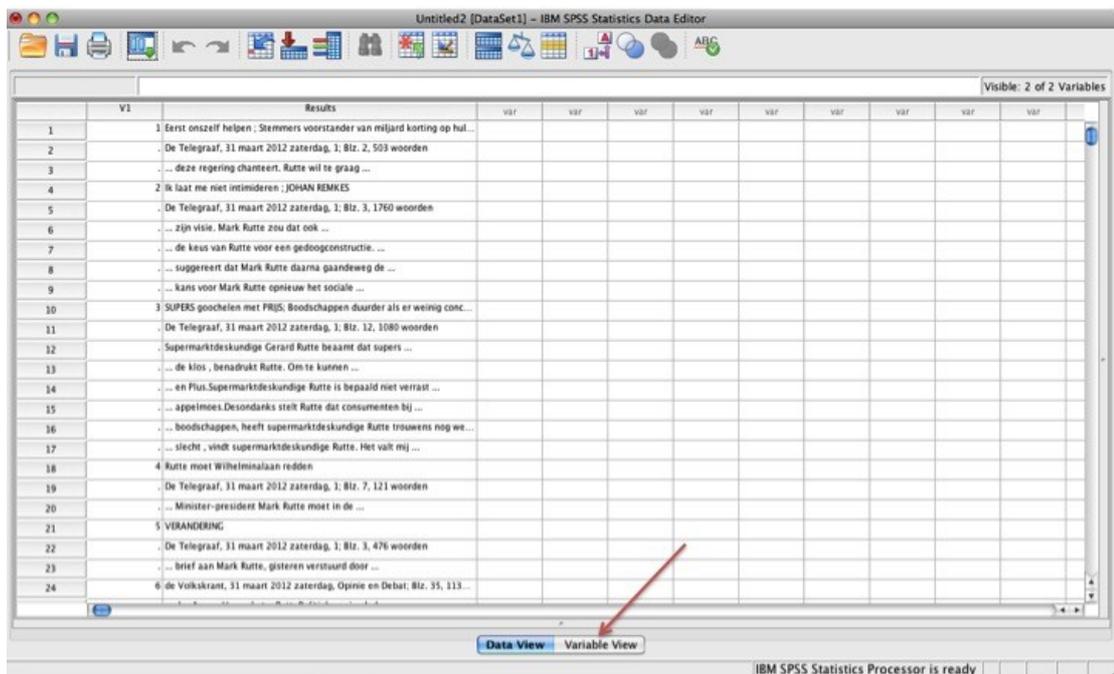
(6) If you have more than one file, please repeat step 3 for all files. Make sure that the "width" of the Results variable is similar for all files – you can adjust that manually it in the 'variable view' screen (see arrow in picture above). You can add the files with the syntax. Please check whether your variable is indeed called 'Results' (and not, for example, 'Resultaten'). If it is called differently, change it by going to 'variable view' (see arrow in picture above) and adjusting the name manually.

(7) If you want to know the number of times (part of) your search string is present in the articles, you have tell spss to look for those terms as well. The syntax creates a variable that gets a '0' if the respective word is mentioned and a '1' if it's not.

(8) The next step is to extract the newpaper in which the article was published. Other sources can be added to the list. The source will be stored in the variable 'newspaper'.

(9) The next step will be to identify the date of the article. Several variables are created, of which four are of relevance: date, yq (quarter in which the article is published), ym (month in which it is published) and yw (week in which it is published). NOTE: you might get some error messages in the output – please ignore them.

(10)   The page number will be extracted and will be stored in the variable 'pagenumer'.

(11)   The number of words will be extracted and will be stored in the variable 'words'.

(12)   The final step 'aggregates' the data. This means that a new file is constructed, that includes one line for each article. All relevant meta-information (source, date, pagenumber, number of words) and also the variable(s) that include the search terms. Please make sure that you add those to the aggregation command.

(13)   Open the new file and you're ready to go to make all kind of analyses. In the syntax, an example for a viability graph is included.