5

Issuecrawling

Building lists of URLs and mapping website networks

Richard Rogers

Introduction: making URL lists of right-wing populist and extremist groupings

According to claims made by the popular press and the think tank Demos we are witnessing the rise of a new kind of populist politics defined by opposition 'to immigration and concern for protecting national and European culture, sometimes using the language of human rights and freedom' (Bartlett, Birdwell and Littler 2011). This 'new right' movement is said to be supplanting the (fascist or neo-Nazi) old guard in a series of European countries, with an orientation distinctive from the 'blood and soil' pathos of old (Van Gilder Cooke, 2011). This chapter describes how we might examine these claims empirically through an online, interdisciplinary approach that combines crawling techniques from web science and close reading of websites from media studies.

The 'how to' research protocol that follows describes how to build lists of URLs to seed link crawling software and ultimately make link maps of right-wing extremism and 'new right' populism in particular European countries. The maps show links between websites, or online networks of websites that can be analysed according to a series of technical characteristics, but here a substantive analysis is also undertaken to examine the claims made. These methods may be situated alongside reading party manifestos and favoured literature, going native by embedding oneself in the groups, interviewing imprisoned or former group members, and other qualitative techniques to distil significant content. The online mapping method of issuecrawling can thus be considered either as an exploratory step that provides leads for further in-depth analysis, or as a means to create country reports with a broad stroke, as is the intention of the longer analysis behind this piece (Rogers 2013).

The exercise commences with the collection of the URLs of populist right-wing and right-wing extremist websites in a series of countries named in popular press articles as well as the Demos study: Austria, Belgium, Bulgaria, Denmark, France, Germany, Greece, Hungary, Italy, the Netherlands, Norway, Portugal, Romania, Serbia and Spain. Lists of websites are made by following a heuristic known as the 'associative query-snowballing technique'. (For a step-by-step elaboration, see the research protocol below.) Queries are formulated, and made in the local domain Googles associated with the countries in question (such as google.at, google.be, google.

bg and google.rs), in the respective local languages, largely in these styles: [populist right parties] as well as [right-wing extremist groups]. When the names of parties, groups or other related entities (e.g. a webshop selling right-wing t-shirts, music and literature) are found, they are entered as lists (each in quotation marks) into the search boxes of the respective local domain Googles, and the results are read. This process is repeated, until no new names are found. That is, lists of populist right and extremist groups are slowly built up from query results. Once the lists gathered from the web search engines are finished, they are compared to expert lists. To find expert lists, queries are made in Google Scholar, first in the home language, and sub-sequently in English. The queries made are similar to those entered in the local domain Googles, in the first round of list-building from the web. Any new groups found on the expert lists in the scholarly literature are searched for online, and if they have a web presence, they are added. Thus, the expert lists add to the web lists. For each group, actor or entity on the list there should be an accompanying URL or multiple URLs.

The work of locating URLs might be arduous for the new right's web presence could be 'on the move', dodging authorities, as is the case in many countries such as Germany where website owners regularly change URLs (and hosts to outside the country) and also move to social media such as Facebook, so as to attract a larger following and make it more burdensome for the authorities to take down what it construes, nationally, as 'illegal content' (Prodhan and Lauer 2016).

Quanti-quali analysis of European right-wing formations online

The URLs of the populist right, the extreme right and the populist and extreme right together are crawled, per country, in three, separate analytical procedures, using the Issuecrawler (issuecrawler.net). Of interest are the comparative sizes of the populist and extreme right as well as other indicators of activity such as responsiveness and freshness. By responsiveness is meant whether the sites are online, and return a response code (or http status code) of 200, when loaded in a browser. Freshness concerns its last update, and its recent consistency in updating.

The two seed sets are crawled together, as well, to compare them and gauge their intercon-nectedness. Do they form one cluster, or are they (largely) separate? Doing this enables one to begin to examine the claims that the populist right is distinctive (clustered separately) and over-taking the old guard, at least according to online network analysis, including responsiveness and freshness. For the analysis, one asks, does the new right have larger, denser clusters and more active and fresher websites than those of the old guard? In most countries under study the answers are in the affirmative, thus largely confirming the popular press and think tank claims.

In terms of the method, for each set of populist, extreme right and combination of URLs, automated co-link analysis is performed, with 'privileged starting points' (a special setting), keeping the seeds on the map, if linked to, whereby those websites receiving at least one link from the seeds are retained in the network. 'Newly discovered' sites are required to receive two links to be included in the network (standard 'co-links'). The 'privilege starting points' feature gives the seeds an increased chance of remaining in the network.

Each of the networks is visualized as a cluster graph (according to measures of inlink centrality), and the findings are described. First, are there other (heretofore) undiscovered groups found through the link analysis? Co-link mapping is a procedure that discovers related URLs through interlinking. In the event, we found Facebook to be a large node in many countries, which not only is in keeping with the impression of groups 'on the move' to social media but also prompts the question of its (separate) analysis, for Facebook cannot be crawled as above. (Only links to Facebook are on the map, not outlinks from Facebook.) Second, which sites are responsive and fresh? Are they mainly the populist ones? Indeed, the old guard's web in a variety of European

countries is often stale. It also might be of interest to inquire into where the websites are registered and by whom. Are they registered under aliases and hosted outside the country? Or are they registered in country, under one's own names? In certain countries, these are signs that groups are in hiding or operating in plain sight, so to speak. In Germany, the groups often mask themselves, while in Austria they tend to operate out in the open.

Apart from the 'technical' characteristics of the websites in the networks (network size and density as well as site responsiveness, freshness, geo-registration and use of alias) the qualitative analysis we conducted concerns the groups' orientation as well as activities, especially in their outreach, forms of communication as well as youth recruitment. Is there an active music scene? Where does one go to participate in person in populist and extreme right-wing culture? Generally, the substantive characteristics of the right-wing formations online specific to the country may be understood by spending significant analytical time reading the websites on the (clickable) Issuecrawler map of each of the national right-wing scenes in question, picking out significant themes, which vary from country to country. In Hungary, for example, the supposed Mongolian language roots have been appropriated by the right (old and new), and the question might be asked: how to take back the yurt. Unlike in Bulgaria (and Spain), where the old guard still thrives (online), in Serbia there is a new, right-wing civil society, with think tanks, which seek to shape the discussion on the future of Serbia around the questions of land and Kosovo. France, witnessing the rise of identitarian (youth) groups and ethno-differentialism, is a dividing line between northern and southern Europe in the sense that counter-jihadism (also referred to as anti-Islam and Islamophobia) is present but not a dominant theme in the new populist right. In Denmark, Norway and to an extent the Netherlands, counter-jihadism increasingly organizes the new right, and indeed here we find especially some of the language of the new right the London think tank described. The claim that the new right employs a vocabulary of immigration opposition borrowed from 'rights talk' is difficult to pinpoint, but the broader claim can be nuanced through the observation that the new right in question is geographically distinctive, and located in northern Europe. In Austria, contrariwise, the populist right's is an anti-capitalist critique (against lavish Austrian balls, and the storage of Austrian gold abroad). In Germany, there is (still) a preponderance of 'brown culture'.

In the following the list-building technique is elaborated in more detail, prior to a reflection on the types of lists that may be authored with the aid of search engines these days, now that the editorial practice of creating web directories has waned.

Research protocol: URL list-making with the associative query-snowballing technique

The objective is to assemble three URL lists per country under study: extreme right, populist right and a combined list. 'Extreme right' and 'populist right' are broad terms not categorized in advance, but instead the authors of online lists classify them as such.

Below are the step-by-step instructions on how to make a list through what is termed associative query-snowballing. The example of list-building is for the 'extreme right' in Spain, however, the process is much the same for any country. The third list is made eventually by merging the first two.

Part I: Making a URL list using the technique

1   Load the local domain Google search engine for the country in question in the browser, e.g. google.es. Design a broad query that will output extreme right groups in Spain. For example, we used: 'Grupos de Extrema derecha en España' (translation: 'Extreme right groups in Spain').
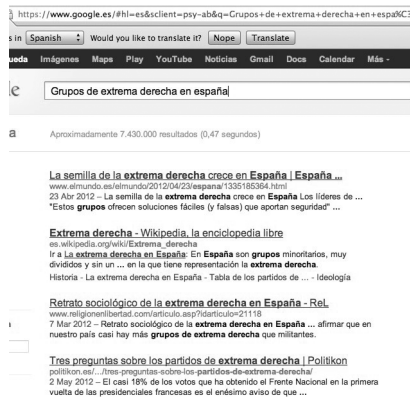
Figure 3.5.1   Google.es results of a query for right-wing extremist groups in Spain. Screenshot, 4 September 2012.

Figure 3.5.2   Simple spreadsheet with names of groups and URLs per group. Screenshot, 4 September 2012.

2   After performing the query, the user is returned a set of results, some of which are lists. List is meant in a broad sense. For example, a news article that reviews the most influential extreme right-wing groups usually will name a number of them. One might find that the article refers to parties or groups not only from the country in question but also to other international groupings. From the pages and articles, the researcher needs to extract the names of the groups that correspond to the country in question, and also find the URLs and include them in a spreadsheet. Let us say in this first step two main groups have been found: España 2000 and Plataforma per Catalyuña (see Figure 3.5.2).

3   Return to Google.es. Enter the names of the groups found in the previous search results as a query using quotation marks: ['España 2000' 'Plataforma Catalunya']. The fresh set of results returned contain ideally not only the two groups used in the query but also new ones that will be associated with them (associative snowballing). Comb through the results, select the names of the new groups and add them to the spreadsheet. For example, the first result contains the new name, 'Democracia Nacional' (see Figure 3.5.3).

4   Enter the two initial groups ('España 2000'and 'Plataforma per Catalunya') together with the new group ('Democracia Nacional') in the search box. Again, one will receive results in which the three groups may be associated with other groups. Add the new ones, including their URLs, to the spreadsheet.

5   Repeat until no new groups are found. For the purposes of robustness one might wish to make queries that contain new combinations of fewer groups.

6   As a note, the last groups to make the lists could be thought of as marginal or historical. It is advisable, as a last step, to query the marginal groups separately, which ideally will return a new set of even more marginal groups, though these also could be from other countries. Repeat until no new country-specific results are found.

Figure 3.5.3   Associative query-snowballing technique, second iteration. Results of the Google.
es query for Plataforma Catalunya and España 2000 yield a third group, Democracia Nacional,
which is then added to the spreadsheet, with its URL. Screenshots, 4 September 2012.

Part II: Finding expert lists, compiling them, adding them to the web list, and making the final list (the web + expert list)

1   Search for academic literature that mentions the extreme right in Spain. Academic articles
and grey literature case studies usually have their own collections of names. One may use
Google Scholar to query in the original language or in English, again employing the broad
search terms: [extreme right–wing Spain]. From the results explore and choose approximately
three or more articles that you have detected containing lists. Recall that lists do not always
look like lists.

2   Extract the names of the groups, and search for the groups' URLs, if (as is often the case)
they are not included. Make a list of all groups and URLs. This is the expert list.

3   Compare the web list (from the associative query–snowballing technique) with the expert
list. There is a list comparison tool, 'triangulation' at https://tools.digitalmethods.net/beta/
triangulate/. It shows the URLs unique to each list as well as those that are common.

4   Take note of the groups or other entities that are unique to the expert list or to the web list. Query the unique groups' names in the search engine, and ascertain whether it has one or more URLs. Retain those groups on the expert lists that have a web presence, i.e. one or more associated URLs claiming to represent or give significant voice to the group.

5   Concatenate the URLs from the web list and the expert list.

Finally, one may take note of what the web yields in comparison to the experts. One may compare epistemologies (how lists are made) as well as ontologies (types of lists). Expert lists (including Wikipedia's) are often exhaustive and alphabetical, and include historical actors, while web lists outputted by search engines are, in the main, hierarchical and fresh.

Conclusion: web and expert URL lists

List-building in preparation for seeding the Issuecrawler or other link crawling software such as Hyphe or VOSON often relies on 'link lists' (Jacomy, Girard, Ooghe-Tabanou and Venturini 2016; Ackland *et al.* 2006). In the past preferred starting points were those lists maintained by Dmoz.org, the open directory project, and Yahoo!, the original web 'directory'. Both projects are dormant. To a degree, directories of all kinds on the web have been supplanted by search engines, which also author lists, albeit of query results rather than list of websites categorized by human editors. Inter-governmental organizations as well as NGOs also have been keepers of expert lists, but their curation practices (such as Amnesty International's list of human rights organizations) have been in abeyance for years. Wikipedia continues to be one of the few human–edited list-makers; given their encyclopaedic quality they require subject–matter expert paring.

The list–making and query-building technique introduced above is designed for a post-directory web. It strives to build lists anew, with the aid of search engines, first by locating lists of mentions of groups, actors or entities (in this case of the right wing), and subsequently by sourcing their URLs, again via search. It is a digital method dubbed 'associative query-snowballing' because each of the actors found has been acquired by association to other actors through iterations of query results.

Acknowledgements

References

Ackland, R., O'Neil, M., Standish, R. and Buchhorn, M. (2006). VOSON: A web services approach for facilitating research into online networks. Paper presented at the *Second International e-Social Science Conference*, 28–30 June, Manchester, UK.

Bartlett, J., Birdwell, J. and Littler, M. (2011). The rise of populism in Europe can be traced through online behaviour . . .: The New Face of Digital Populism: Lega Nord., London: Demos. Retrieved from: http://www.demos.co.uk/files/Demos_OSIPOP_Book-web_03.pdf.

Jacomy, M., Girard, P., Ooghe-Tabanou, B. and Venturini, T. (2016). Hyphe, a curation-oriented approach to web crawling for the social sciences. *Proceedings of the International AAAI Conference on Web and Social Media* (ICWSM-16), Cologne, Germany.

Prodhan, G. and Lauer, K. (25 February, 2016). Germans talk tough, fete Facebook's Zuckerberg, *Reuters*. Retrieved from: www.reuters.com/article/us-facebook-germany-zuckerberg-idUSKCN0VY2DD

Rogers, R. (2013). Right-wing formations in Europe and their counter-measures: an online mapping. Amsterdam: Govcom.org Foundation and the Digital Methods Initiative. Retrieved from: https://wiki.digitalmethods.net/Dmi/RightWingPopulismStudy.

Van Gilder Cooke, S. (29 July, 2011). Europe's right wing: a nation-by-nation guide to political parties and extremist groups. *Time Magazine*. Retrieved from: http://content.time.com/time/specials/packages/article/0,28804,2085728_2085727_2085712,00.html