

Notes on `case_when` in `10_text_prep_FB.R`

Brian Brew

1/17/2022

Senate Notes

Some examples of the sorts of links that are included in non-classified texts:

- Campaign websites (i.e. amymcgrath.com, hickenlooper.com, paulajeane.com, etc.)
- What I take to be campaign contribution prompts on campaign websites (i.e. act.amymcgrath.com, action.hickenlooper.com, secure.hickenlooper.com, go.dralgrossak.com, secure.chriscoons.com, act.jonipac.com, go.jaimeharrison.com, etc.)
- Links to state party websites (i.e. coloradodems.org, texasgop.org)
- Government websites (i.e. absentee ballot request/voting information pages, like vrsws.sos.ky.gov, sos.la.gov)
- Lots of news sources – the trouble with these is, there’s a whole bunch of them – it might not be so feasible to code them all, not least because there are so many variations (not just NYT/WaPo/CNN/Fox, or growing online outlets like Vox/Politico/The Hill, but also a host of outlets like the Dover Post, the Alaska Public, the Sioux City Journal, and many, many more)
- Some expired sites that I’m not 100% sure what their goal is – (i.e. one called getthefacts.co [probably an independent partisan source, maybe a fact-checker], another called jhfp.cp.bsd.net [who knows])
- Fact-checkers (i.e. politifact.com)
- Links to forums – John Hickenlooper linked what looks like a connection to send him a Facebook message in one ad (it shows up as fb.com...)
- Links to mobilization efforts/voter information (i.e. events.mobilizeamerica.io, and <https://www.mobilize.us/>, which both lead to the same page; votebluesc.com, which is partisan; votemaine.com, which looks nonpartisan [which is in-character for Maine])
- victoryaction.com on some Republican pages... could be, perhaps, the National Victory Action Fund? Cory Gardner, former CO senator, is its chair... not the most Trump-y Republican in recent memory by any stretch...
- Forums/Think Tanks (i.e. justsecurity.org)
- Attack sites against their opponents (i.e. johnjamesrevealed.com, grosstapes.com [referring to Al Gross], massachusettsmanny.com [referring to Bill Hagerty (TN)’s MA-born primary challenger, Manny Sethi], thetruthabouttheresa.com, lindseyhaschanged.com, etc.)
- Links to other politicians (Chris Coons of Delaware has a link to joebiden.com)
- The suffix `.bsd.net` appears on expired links for multiple candidates...

```

# The current case_when() code:
type = case_when(
  grepl("actblue.com", ad_creative_link_caption) ~ "ActBlue",
  grepl("winred.com", ad_creative_link_caption) ~ "WinRed",
  grepl("ngpvan.com|myngp.com", ad_creative_link_caption) ~ "NGP VAN",
  grepl("anedot.com", ad_creative_link_caption) ~ "Anedot",
  grepl("victorypassport.com", ad_creative_link_caption) ~
    "Victory Passport",
  grepl("fundraiser", ad_creative_link_caption) ~ "Misc.",
  is.na(ad_creative_link_caption) ~ "Non-financial",
  # grepl(" ", ad_creative_link_caption) ~ "Non-financial",
  grepl(
    "conversation with |town hall|meet |tour stop |iwillvote.com",
    ad_creative_link_caption
  ) ~ "Non-financial"
),
# Possibly relevant additional grepl lines:
  grepl(".gov", ad_creative_link_caption) ~ "Governmental Information",
# If there's a way to have some exclusionary sides to grepl, could get some good
# use out of that. I know there's a lot of act.actblue.com, secure.winred.com,
# etc. -- if we could have a case when after we've isolated ActBlue, WinRed,
# etc., we could then do a grepl that could potentially identify candidate
# contribution webpages, something along the lines of:
## grepl("secure.|act.|action.|go", ad_creative_link_caption) ~ "Personal
### Contribution Link",
# There would be a few logistical hiccups to sort out (i.e. other links that
# could have the same sorts of prefixes), but it's worth considering

## Is there a way to use grep/grepl to pull a chunk of text for another column?
## If there is -- for instance, if we could pull the last word, like MCGRATH,
## from the `candidate` column of fb_unique[["senate"]], and set it tolower()
## and call it something along the lines of "candidate_surname" or some such.
### grepl(candidate_surname, ad_creative_link_caption) ~ "Personal Webpage",
## could be a pertinent line to include.
## I'll do a bit more digging into grepl() to see if this is feasible.
## Of course, folks who don't use their last names in their personal websites,
## like Paula Jean Swearngen of WV (paulajeane.com), would complicate matters...
## Would also need to exclude any campaign-contribution-specific stuff, i.e.,
## wouldn't want corybooker.com and act.corybooker.com to go to the same
## categorization.

# Ads that contain 'vote' might be worth categorizing as non-financial...
# Presumably aimed at voting electorate, not so much the financial electorate

```

House Notes

Overall, things look pretty darn similar in terms of the categories between the House and the Senate. Given the volume of ads, I haven't been able to inspect the House as closely as the Senate, but the same general subjects and labels strike me as relevant to the House Facebook Ad.