

Robert Genega

CSMC 471 Project 4

TF-IDF

I ran tf-idf over a subset of the articles in the “articles1.csv” file provided (approximately 200). I found that the top five #1 topics for this dataset were:

1. Republicans
2. Ethics
3. Turkey
4. Cuomo
5. Kushner

Here are the top 10 topics for three articles from this dataset:

Trump Appears to Side With Assange Over Intelligence Agencies

- 1: assange
- 2: intelligence
- 3: manigault
- 4: trump
- 5: wikileaks
- 6: pence
- 7: wednesday
- 8: liaison
- 9: russia
- 10: russian

SpaceX Says It’s Ready to Launch Rockets Again

- 1: spacex
- 2: oxygen
- 3: helium
- 4: launch

5: satellites

6: liquid

7: rocket

8: iridium

9: aluminum

10: carbon

A New Casualty of Syria's War: Drinking Water in Damascus

1: water

2: damascus

3: wash

4: rebels

5: crisis

6: syrian

7: dishes

8: wells

9: malik

10: Aleppo

Word Cloud: House Republicans Fret About Winning Their Health Care Suit

