

Topic Shifts Between Two US Presidential Administrations

**Ziquan Wang
Borui Lin
Ian Milligan
Jimmy Lin**





**Born-Digital Cultural
Heritage is Important for
Understanding History.**



the WHITE HOUSE PRESIDENT BARACK OBAMA

Contact Us Get Email Updates



BRIEFING ROOM

ISSUES

THE ADMINISTRATION

PARTICIPATE

1600 PENN



AN IN-DEPTH LOOK AT PRESIDENT OBAMA'



INSIDE THE WHITE HOUSE

[West Wing Tour](#)
[Eisenhower Executive Office Building Tour](#)
[Video Series](#)
[Décor and Art](#)
[Holidays](#)
[See All](#)

HISTORY & GROUNDS

[Presidents](#)
[First Ladies](#)
[The Vice President's Residence & Office](#)
[Eisenhower Executive Office Building](#)
[Camp David](#)
[Air Force One](#)

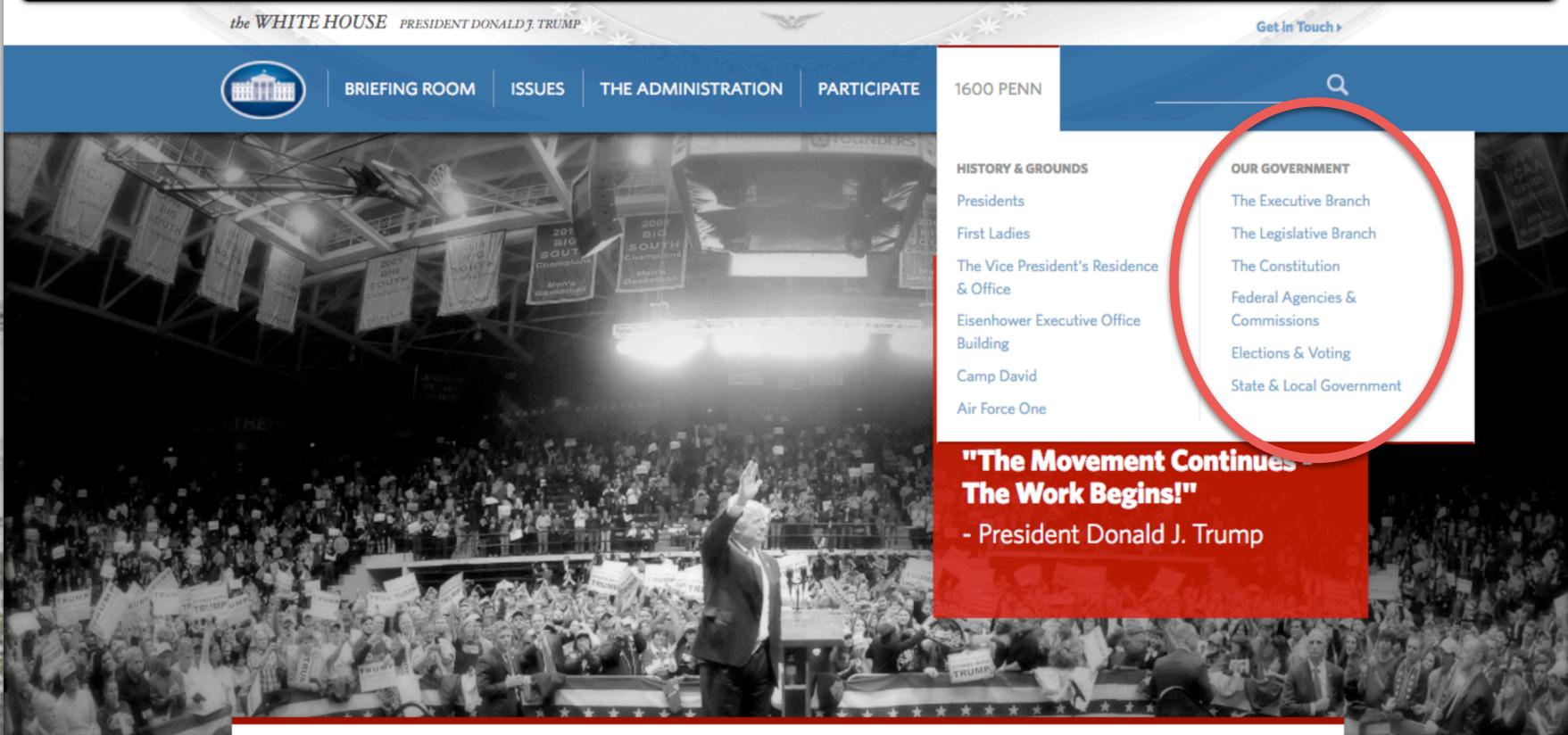
OUR GOVERNMENT

[The Executive Branch](#)
[The Legislative Branch](#)
[The Judicial Branch](#)
[The Constitution](#)
[Federal Agencies & Commissions](#)
[Elections & Voting](#)
[State & Local Government](#)
[Resources](#)



THE ADMINISTRATION

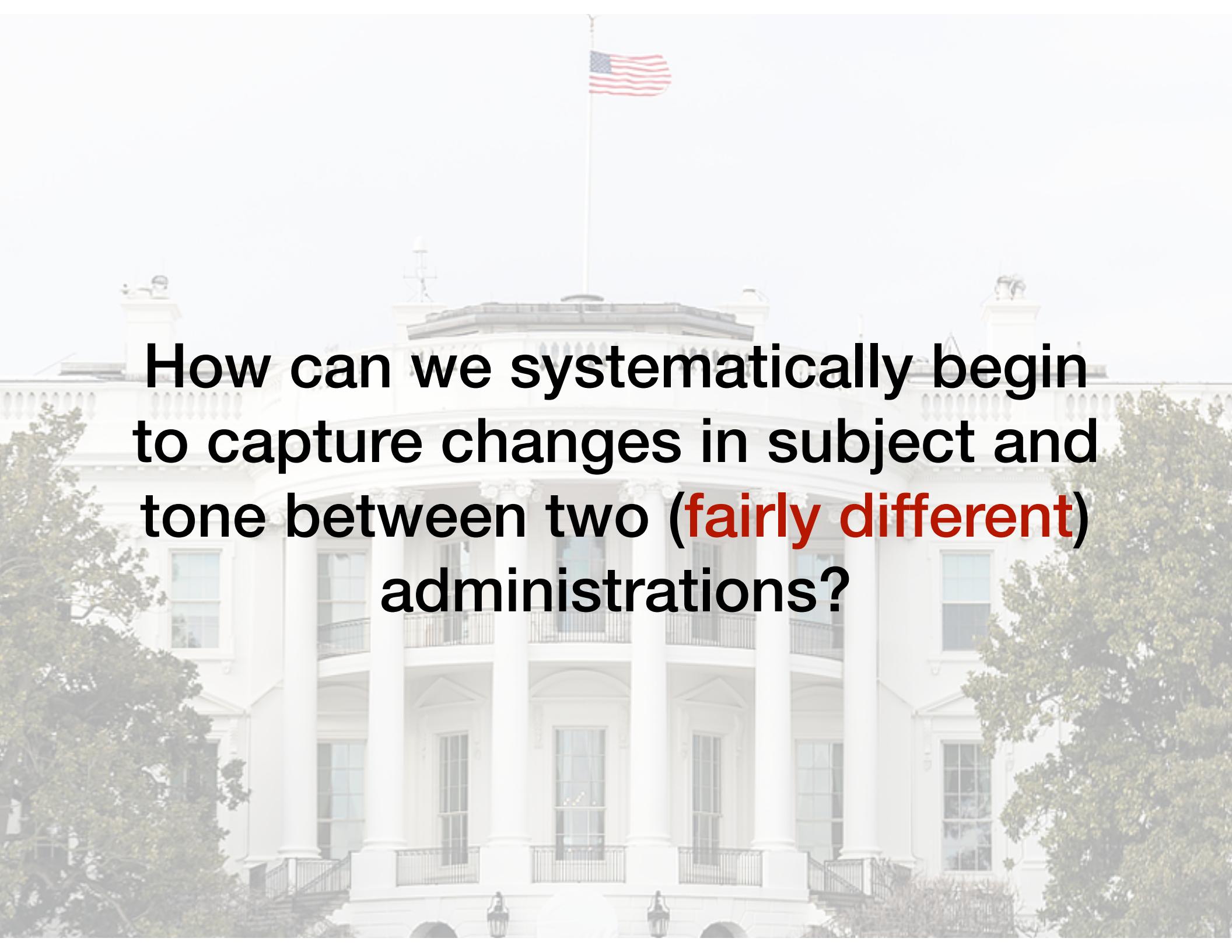




The Inauguration of Donald J. Trump

[WATCH THE VIDEO](#)





How can we systematically begin
to capture changes in subject and
tone between two (**fairly different**)
administrations?

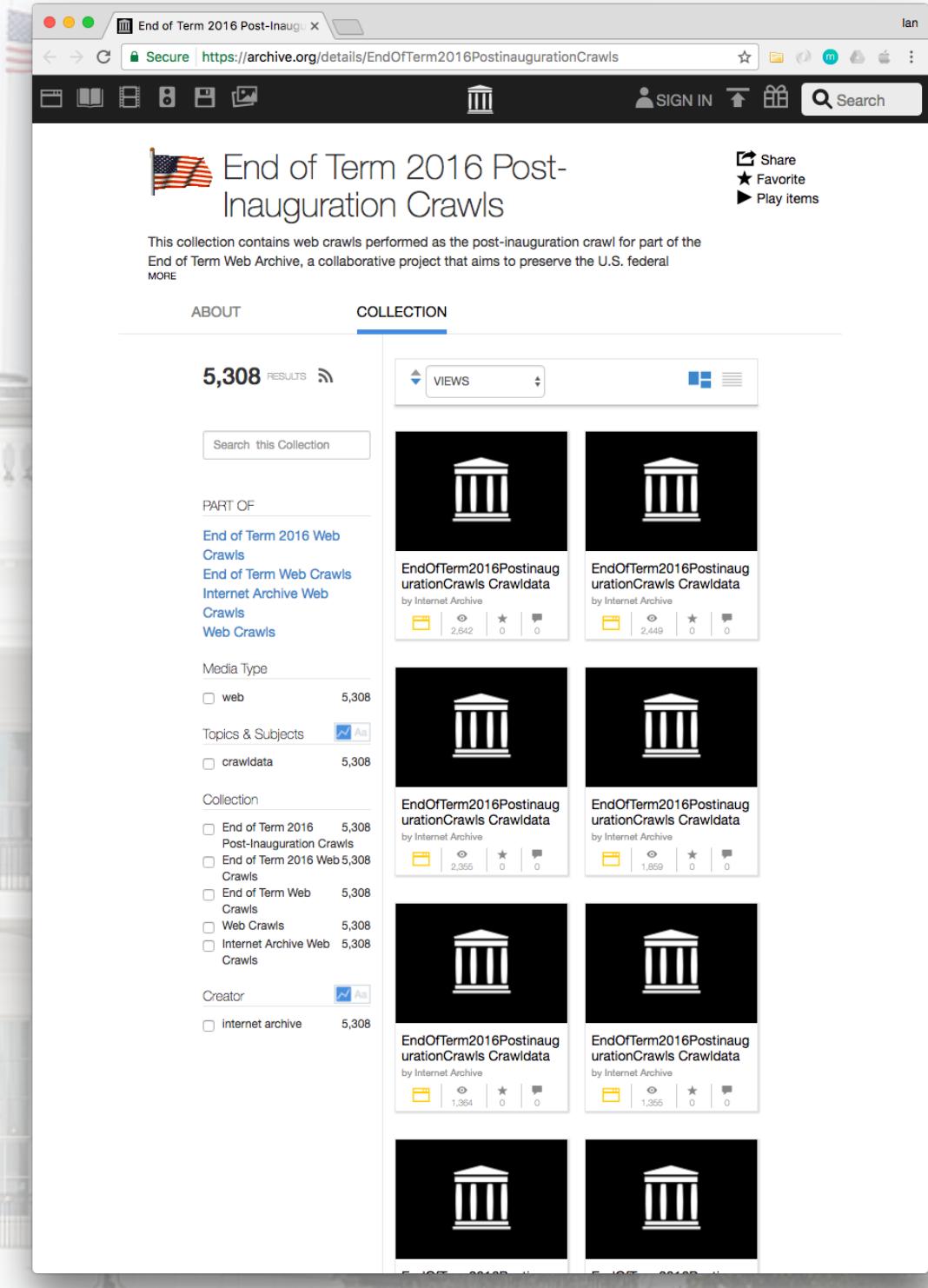


**To help move beyond
anecdote and towards
distant reading.**

Archive-It Datasets.

Obama (320 GB)
Trump (13 GB)

No, that's not a rounding
error.



The screenshot shows a web browser window displaying the Internet Archive collection page for the "End of Term 2016 Post-Inauguration Crawls". The title bar reads "End of Term 2016 Post-Inauguration Crawls" and the URL is "https://archive.org/details/EndOfTerm2016PostinaugurationCrawls". The page features a header with a US flag icon, the collection name, and options to "SIGN IN", "Share", "Favorite", and "Play items". Below the header, there are tabs for "ABOUT" and "COLLECTION", with "COLLECTION" being the active tab. It displays "5,308 RESULTS". A search bar says "Search this Collection". On the left, there are filters for "PART OF", "Media Type", "Topics & Subjects", "Collection", and "Creator". The main content area shows a grid of 12 items, each represented by a small thumbnail of a classical building and a title like "EndOfTerm2016PostinaugurationCrawls Crawldata by Internet Archive". Each item has a small preview icon, a file count (e.g., 2,842, 2,449), and a rating section with stars and a comment icon.

Processing using Warcbase

- **Warcbase**: an open-source platform for managing web archives (<http://warcbase.org>)
 - Scriptable analytics and data processing
 - Scalable using Apache Spark
 - Documentation at docs.warcbase.org

The screenshot shows the GitHub repository page for 'lntool/warcbase'. The repository has 713 commits, 11 branches, and 0 releases. The README.md file is visible, containing the following text:

Warcbase is an open-source platform for managing analyzing web archives <http://warcbase.org/>

Branch: master New pull request Create new file Upload files Find

ianmilligan1 committed on GitHub Updating README.md for link visualizer
vis Updating README.md for link visualizer
warcbase-core better io (#261)
warcbase-hbase Cleaned up and simplified dependencies, etc.
.gitignore Copy dependencies into solr home.
.travis.yml Moved HBase and non-core code out of warcbase-core into warcbase-hbase.
CONTRIBUTING.md draft of contributing file
README.md Minor tweak to README.
pom.xml Upgraded to CDH 5.7.1.

README.md

Warcbase build passing

Warcbase is an open-source platform for managing web archives built on Hadoop and HBase. The platform provides a flexible data model for storing and managing raw content as well as metadata and extracted knowledge. Integration with Hadoop provides powerful tools for analytics and data processing via Spark.

There are two main ways of using Warcbase:

- The first and most common is to analyze web archives using `spark`, these functions will be covered in the next section.

Processing using Warcbase

- Extract all text
 - Filter by domain (i.e. all pages in whitehouse.gov)
 - Filter by URL pattern (i.e. all pages in whitehouse.gov/hats)
- Filter out boilerplate
- Filter by date
- Filter by language
- Or any of the above...

```
/*
 * Warcbase: an open-source platform for managing web archives
 *
 * Licensed under the Apache License, Version 2.0 (the "License");
 * you may not use this file except in compliance with the License.
 * You may obtain a copy of the License at
 *
 *     http://www.apache.org/licenses/LICENSE-2.0
 *
 * Unless required by applicable law or agreed to in writing, software
 * distributed under the License is distributed on an "AS IS" BASIS,
 * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
 * See the License for the specific language governing permissions and
 * limitations under the License.
 */
package org.warcbase.spark.matchbox

import java.io.IOException
import org.jsoup.Jsoup

object RemoveHTML {
  def apply(content: String) = {
    try {
      Jsoup.parse(content).text().replaceAll("[\\r\\n]+", " ")
    }
    catch {
      case e: IOException => throw new IOException("Caught exception processing input row ", e)
    }
  }
}
```



**Our question – how can we do
quick, cheap, and rigorous topic
classification on datasets?**

Topic Classification

- Obtain human annotated training data
 - pairs of (sentence, category)
- Train text classifier
- Apply text classifier to unseen text
- Analyze results

**What if we don't have
any annotated data?**



Bootstrapping to the rescue!

Bootstrapped Topic Classification

- Obtain human annotated training data
 - pairs of (sentence, category)
- Train text classifier
- Apply text classifier to unseen text
- Analyze results

Let's pretend!

Bootstrapping via keywords

- Gather training data via keyword matching:
 - If a sentence contains one or more keywords, assume its about a particular topic

Climate

adaptation, alternative energy, anthropoc*, anthropog*, carbon, cfc, clean energy, climate change, climategate, co2, ...

Finance

adjustable-rate mortgage, bailout, bubble, capital requirement, conservatorship, exposure, fannie mae, financial fraud, foreclosure, freddie mac, ...

Health

health care, healthcare, health insurance, pharmaceutical, drug abuse, alcohol abuse, Obamacare, prescription drug, long term care, rehabilitation, ...

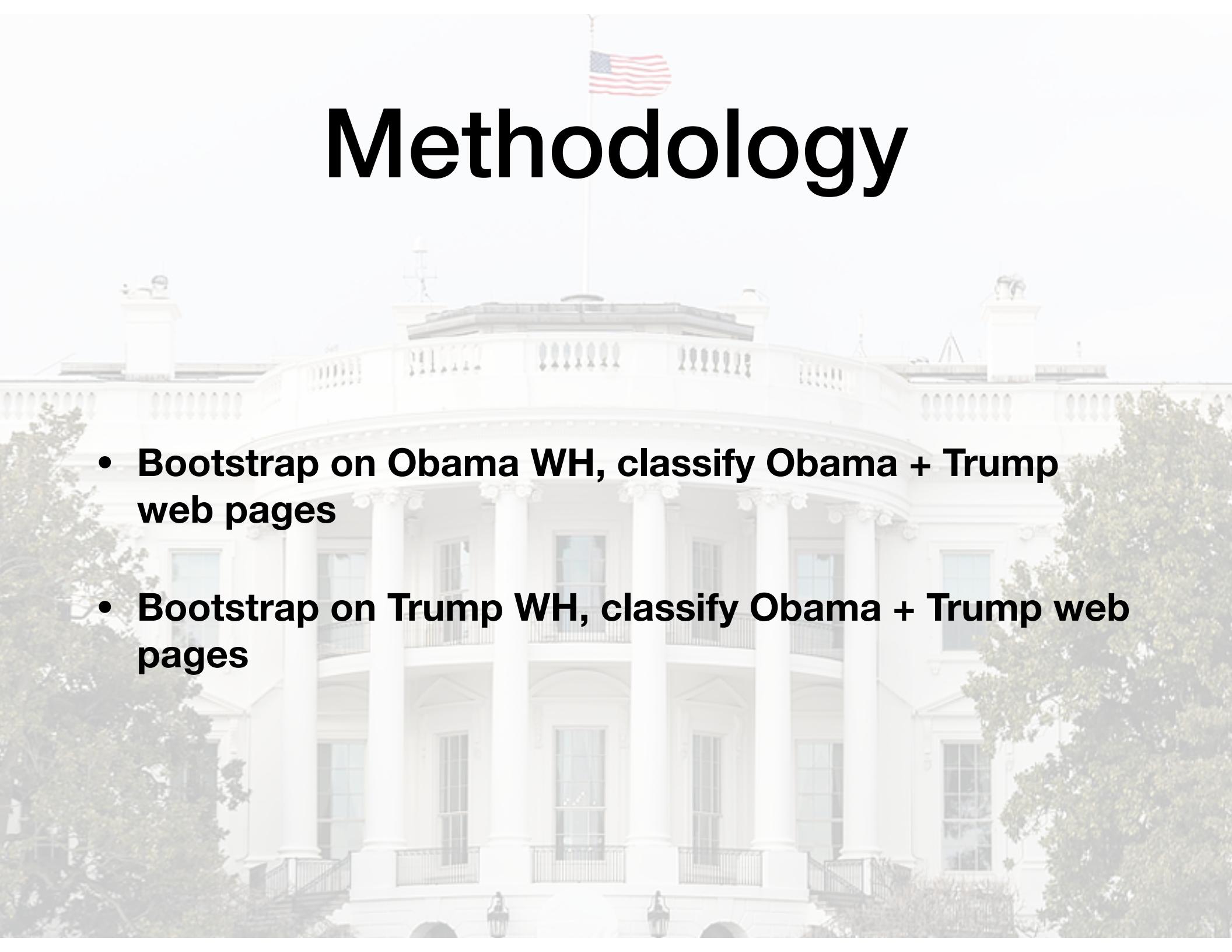
Security

9/11, al-qa*, alien smuggl*, arms prolifer*, arms smuggl*, arms transfer, assassin, atrocit*, authoritarian, ballistic missile, bin laden, ...

Category and keywords from: Emily Gade, John Wilkerson, and Anne Washington.
(2017) The .GOV Internet Archive: A Big Data Resource for Political Science. *The Political Methodologist*.

Why bootstrapping works...

- Start with reasonable (but noisy) labels
- Machine learning algorithm discovers good correlate features
 - i.e. “Obamacare” was a keyword, but not “ACA”
 - In training data, “Obamacare” and “ACA” co-occur
 - Classifier gets the two together, and the second time you run, **a bit more bang for your buck...**



Methodology

- Bootstrap on Obama WH, classify Obama + Trump web pages
- Bootstrap on Trump WH, classify Obama + Trump web pages



Expectations with the data?

Results of the Classifiers on the Trump White House

Classifier trained on Trump

Category	Sentences	Percentage
Uncategorized	1462	95.62
Security	40	2.62
Health	23	1.5
Climate	3	0.2
Finance	1	0.07

Classifier trained on Obama

Category	Sentences	Percentage
Uncategorized	1470	96.2
Security	36	2.4
Health	17	1.11
Climate	5	0.3
Finance	0	0

Results of the Classifiers on the Obama White House

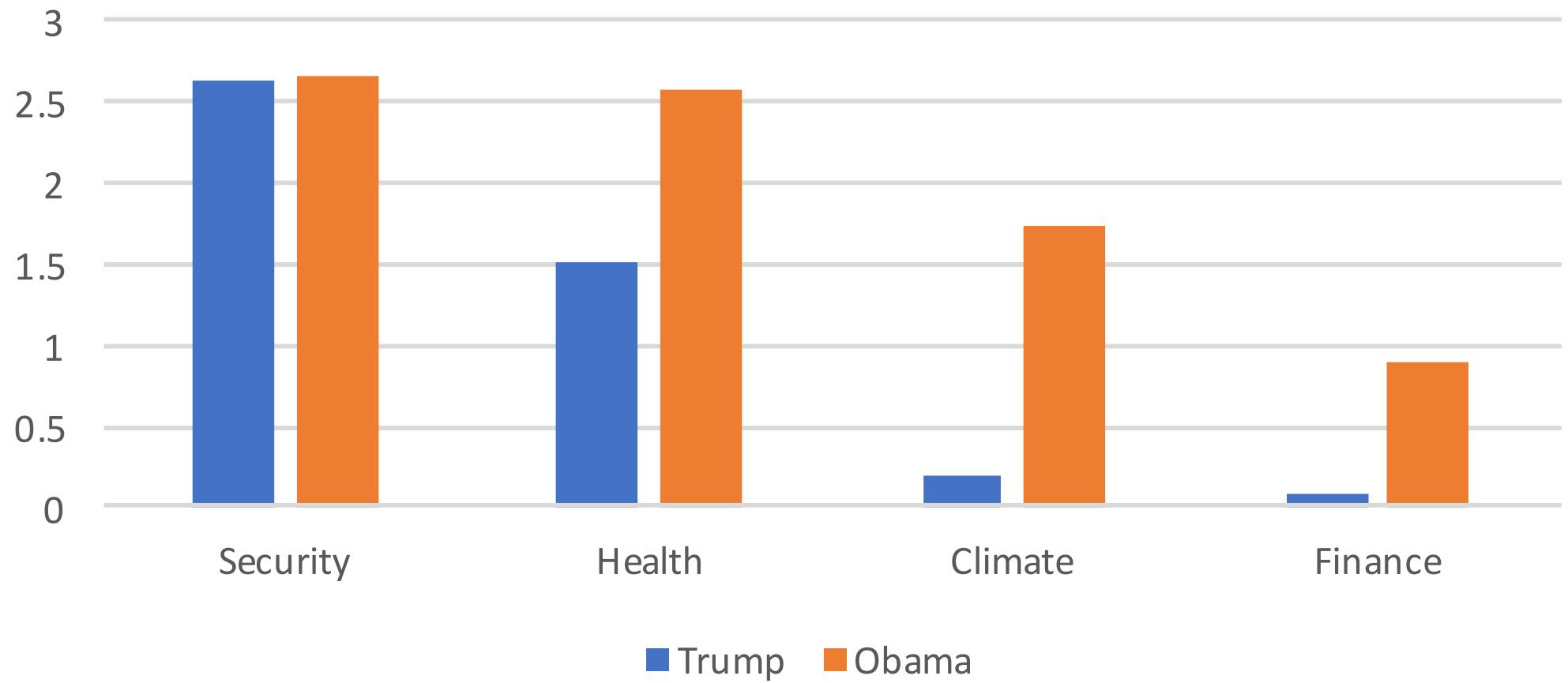
Classifier trained on Trump

Category	Sentences	Percentage
Uncategorized	95975	93.1
Health	2734	2.65
Security	2650	2.57
Climate	1809	1.72
Finance	916	0.89

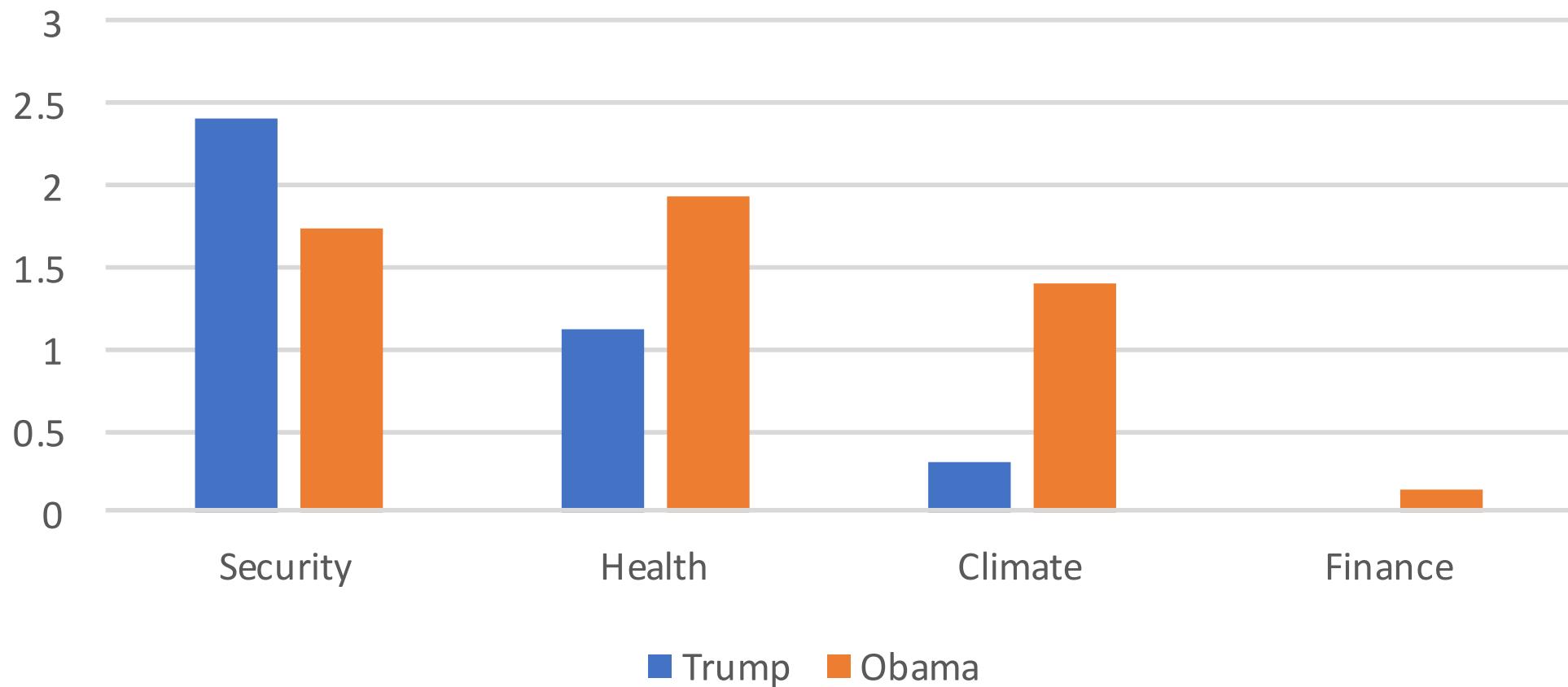
Classifier trained on Obama

Category	Sentences	Percentage
Uncategorized	97381	94.8
Health	1974	1.92
Security	1768	1.72
Climate	1443	1.4
Finance	156	0.15

Comparing WhiteHouse.gov (using Trump classifier)



Comparing WhiteHouse.gov (using Obama classifier)



Document output:

URL, Topic, Ratio of
relevant sentences, Total
classified sentences

The Data: Obama Climate

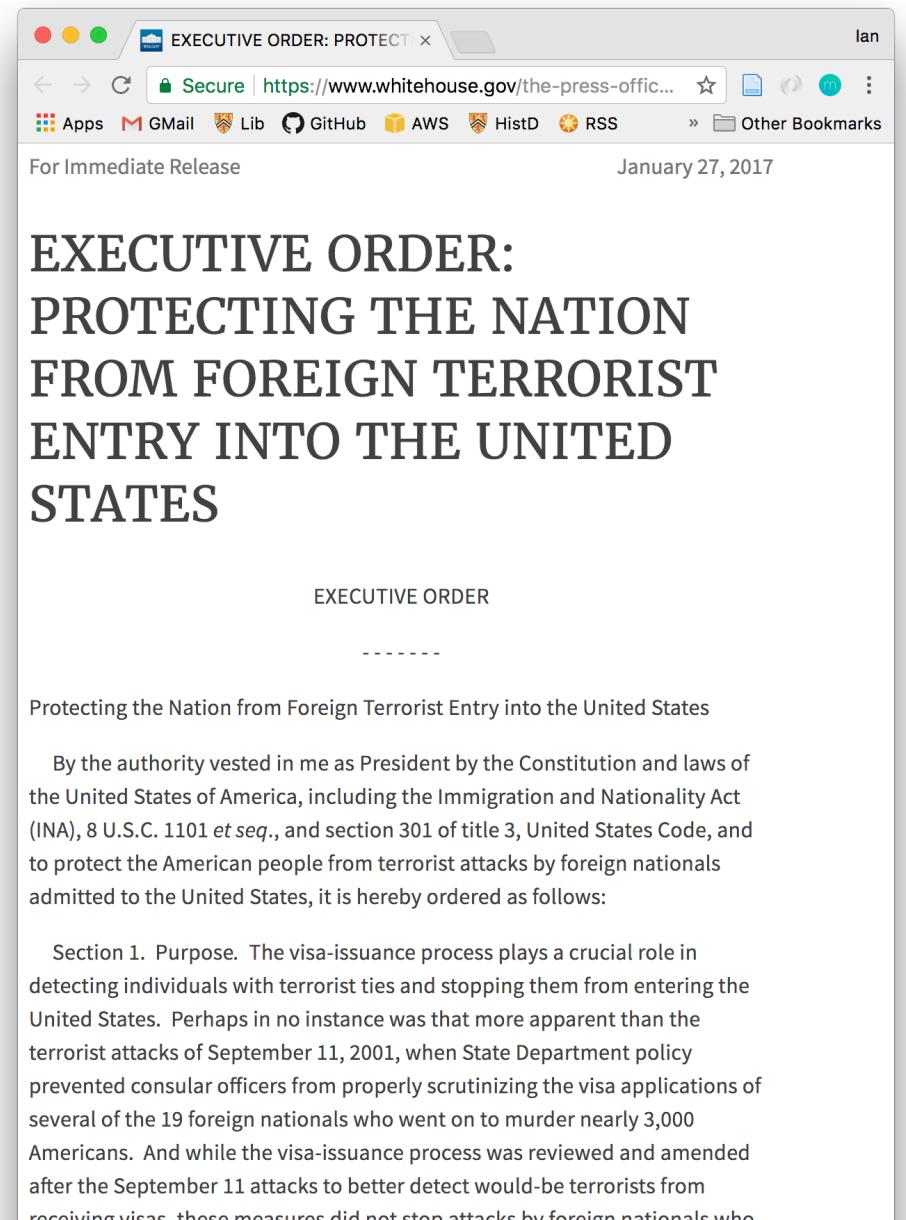
- <https://www.whitehouse.gov/blog/2014/11/12/us-and-china-just-announced-important-new-actions-reduce-carbon-pollution> CLIMATE (0.7619048, 21)
- <https://www.whitehouse.gov/the-press-office/2011/09/02/white-house-officials-hold-background-conference-call-regarding-clean-ai> CLIMATE (0.5, 2)
- <https://www.whitehouse.gov/blog/2011/04/15/west-wing-week-open-business> CLIMATE (0.33333334, 6)
- <https://www.whitehouse.gov/blog/2010/03/15/so-you-want-boost-exports-have-i-got-a-program-you> CLIMATE (0.3181818, 22)
- <https://www.whitehouse.gov/the-press-office/2016/02/10/statement-press-secretary> CLIMATE (0.4, 5)
- <https://www.whitehouse.gov/the-press-office/2015/11/16/new-report-provides-authoritative-assessment-national-regional-impacts> CLIMATE (0.45945945, 37)
- <https://www.whitehouse.gov/blog/2011/10/07/link-between-american-energy-and-prosperity> CLIMATE (0.3809524, 21)
- <https://www.whitehouse.gov/photos-and-video/video/clean-energy-economy-forum-one> CLIMATE (1.0, 1)
- <https://www.whitehouse.gov/the-press-office/2015/04/30/statement-press-secretary-s-535> CLIMATE (1.0, 1)
- <https://www.whitehouse.gov/blog/2013/07/17/preparing-new-york-climate-and-health-impacts> CLIMATE (0.45833334, 24)
- <https://11111011100.whitehouse.gov/sotu/climate-share> CLIMATE (1.0, 1)
- <https://11111011100.whitehouse.gov/sotu/1399fbda5ca8ee0cced4d6394c4b7ea> CLIMATE (1.0, 1)
- <https://11111011100.whitehouse.gov/energy> CLIMATE (1.0, 1)
- <https://www.whitehouse.gov/blog/2013/06/29/weekly-address-confronting-growing-threat-climate-change> CLIMATE (1.0, 4)
- <https://11111011100.whitehouse.gov/president-obama-climate-action-plan> CLIMATE (0.62666667, 75)
- <https://11111011100.whitehouse.gov/administration/eop/cea> CLIMATE (0.5, 4)
- <https://budget2017.whitehouse.gov/template/directives/navigation/breadcrumbs.html> CLIMATE (1.0, 1)

The Data: Trump Security

- <https://www.whitehouse.gov/trump-stands-with-israel> SECURITY (0.33333334, 3)
- <https://www.whitehouse.gov/the-press-office/2017/02/17/readout-presidents-call-president-beji-caid-essebsi-tunisia> SECURITY (0.75, 4)
- <https://www.whitehouse.gov/the-press-office/2017/01/27/executive-order-protecting-nation-foreign-terrorist-entry-united-states> SECURITY (0.38271606, 81)
- <https://www.whitehouse.gov/the-press-office/2017/02/13/readout-vice-presidents-call-president-michel-temer-brazil> SECURITY (0.33333334, 3)
- <https://www.whitehouse.gov/the-press-office/2017/02/13/readout-vice-presidents-meeting-canadian-ministers> SECURITY (0.5, 6)
- <https://www.whitehouse.gov/trump-stands-with-israel> SECURITY (0.33333334, 3)
- <https://www.whitehouse.gov/featured-videos/video/2017/02/20/president-trump-announces-hr-mcmaster-national-security-advisor> SECURITY (1.0, 1)
- <https://www.whitehouse.gov/the-press-office/2017/02/19/readout-presidents-call-prime-minister-keith-rowley-trinidad-and-tobago> SECURITY (0.33333334, 3)
- <https://www.whitehouse.gov/the-press-office/2017/02/18/readout-vice-presidents-meeting-prime-minister-yildirim-turkey> SECURITY (0.42857143, 7)
- <https://www.whitehouse.gov/the-press-office/2017/02/18/readout-vice-presidents-meeting-presidents-lithuania-latvia-and-estonia> SECURITY (0.5, 4)
- <https://www.whitehouse.gov/the-press-office/2017/01/28/presidential-memorandum-organization-national-security-council-and> SECURITY (0.42105263, 57)
- <https://www.whitehouse.gov/the-press-office/2017/02/17/readout-presidents-call-president-beji-caid-essebsi-tunisia> SECURITY (0.75, 4)

Example: Security

- Close reading let us do some checks to see if things were making sense
- Trump White House, documents with high percentage of security, seen at right – executive order on travel ban.



The screenshot shows a web browser window with the title "EXECUTIVE ORDER: PROTECT". The address bar indicates a secure connection to <https://www.whitehouse.gov/the-press-office/2017/01/27/executive-order-protect-nation-foreign-terrorist-entry-united-states>. The page header includes "For Immediate Release" and the date "January 27, 2017". The main content is titled "EXECUTIVE ORDER: PROTECTING THE NATION FROM FOREIGN TERRORIST ENTRY INTO THE UNITED STATES". Below the title, it is labeled "EXECUTIVE ORDER". A section titled "Protecting the Nation from Foreign Terrorist Entry into the United States" contains the text: "By the authority vested in me as President by the Constitution and laws of the United States of America, including the Immigration and Nationality Act (INA), 8 U.S.C. 1101 et seq., and section 301 of title 3, United States Code, and to protect the American people from terrorist attacks by foreign nationals admitted to the United States, it is hereby ordered as follows:". The text continues with "Section 1. Purpose. The visa-issuance process plays a crucial role in detecting individuals with terrorist ties and stopping them from entering the United States. Perhaps in no instance was that more apparent than the terrorist attacks of September 11, 2001, when State Department policy prevented consular officers from properly scrutinizing the visa applications of several of the 19 foreign nationals who went on to murder nearly 3,000 Americans. And while the visa-issuance process was reviewed and amended after the September 11 attacks to better detect would-be terrorists from receiving visas, these measures did not stop attacks by foreign nationals who".

Example: Security

- Obama White House, documents with high percentage of security, seen at right – NSC statement on Rice meeting with UN Special Envoy for Syria.

The screenshot shows a web browser window with the following details:

- Title Bar:** EXECUTIVE ORDER: PROTECT, Statement by NSC Spokesperson
- Address Bar:** https://obamawhitehouse.archives.gov/the-p... (Secure)
- Toolbar:** Apps, GMail, Lib, GitHub, AWS, HistD, RSS
- Header:** Office of the Press Secretary
- Date:** November 05, 2014
- Title:** Statement by NSC Spokesperson Bernadette Meehan on National Security Advisor Susan E. Rice's Meeting with UN Special Envoy for Syria Staffan de Mistura
- Text:** National Security Advisor Susan E. Rice met today with the United Nations' Special Envoy for Syria Staffan de Mistura. Mr. de Mistura updated Ambassador Rice on his recent regional consultations related to seeking a resolution to the Syria conflict and expanded on his October 30 briefing to the UN Security Council. Ambassador Rice emphasized the United States' support for a political solution to Syria's civil war.

Ambassador Rice underscored the United States' commitment to increasing support for the moderate Syrian opposition, including through a train and equip program. She thanked Mr. de Mistura for the renewed energy he has brought to international efforts to bring an end to the war in Syria and discussed measures that the United States and other partners could take to support his efforts. Ambassador Rice also stressed the United States' continued commitment to providing humanitarian assistance to those affected by the conflict.
- Share Buttons:** Twitter, Facebook, Email
- Footer:** HOME, PRINTING ROOM, ISSUES, THE ADMINISTRATION, BROWSE

Example: Health

- Obama White House, documents with high percentage of health, seen at right – Innovation in Substance Use Disorders and Mental Health Treatment.

Innovation in Substance Use Disorders and Mental Health Treatment

SEPTEMBER 16, 2013 AT 5:47 PM ET BY CAMERON HARDESTY

Summary: Today, the Office of National Drug Control Policy hosted a conference at the White House to highlight innovative technologies for substance abuse treatment, wellness, mental and behavioral health. Supported by the Substance Abuse and Mental Health Administration (SAMHSA), the Office of National Coordinator for Health Information Technology (ONC), and the Office of Science and Technology Policy (OSTP), the conference was kicked off by U.S. Chief Technology Officer Todd Park and Office of National Drug Control Policy Director Gil Kerlikowske.

Today, the Office of National Drug Control Policy hosted a conference at the White House to highlight innovative technologies for substance abuse treatment, wellness, mental and behavioral health. Supported by the [Substance Abuse and Mental Health Administration](#) (SAMHSA), the [Office of National Coordinator for Health Information Technology](#) (ONC), and the [Office of Science and Technology Policy](#) (OSTP), the

Examples: Differences and Commonalities

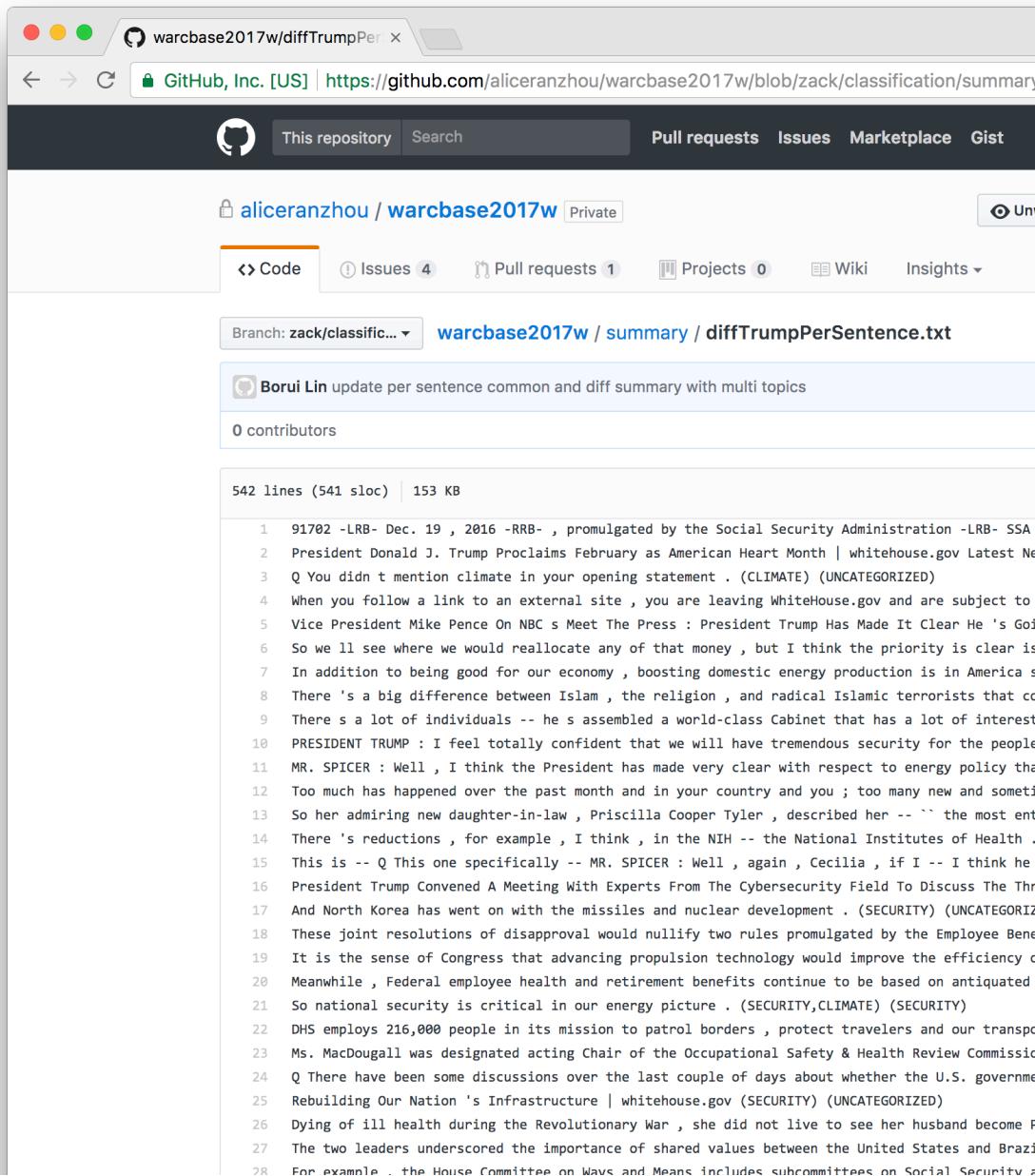
- The method generates documents that had similar detection; but also lets you see the differences.
- For example, URL, TOPIC1, TOPIC2

The screenshot shows a GitHub Gist page titled 'diffObama.txt'. The page displays a list of URLs and their associated topics, comparing them between two classifiers: 'obama' and 'trump'. The URLs are listed in pairs, with the 'obama' classifier's URL on the left and the 'trump' classifier's URL on the right. The topics are listed in brackets next to each URL. The page includes navigation links for Code, Issues (4), Pull requests (1), Projects (0), Wiki, and Insights. The commit history shows a single commit from 'Borui Lin' adding common and diff summary for both classifiers, made 15 hours ago.

URL (obama)	URL (trump)	Topic
https://www.whitehouse.gov/photos-and-video/video/2016/08/20/weekly-address-celebrating-100th-anniversary-national-park-service	https://www.whitehouse.gov/photos-and-video/video/2016/08/20/weekly-address-celebrating-100th-anniversary-national-park-service	UNCATEGORIZED CLIMATE,HEALTH,SECURITY
https://www.whitehouse.gov/photos-and-video/video/2013/03/28/white-house-hangout-maker-movement?page=1	https://www.whitehouse.gov/photos-and-video/video/2013/03/28/white-house-hangout-maker-movement?page=1	UNCATEGORIZED CLIMATE,HEALTH,SECURITY
https://www.whitehouse.gov/photos-and-video/video/2011/05/19/open-questions-president-s-speech-middle-east-and-north-africa?page=1	https://www.whitehouse.gov/photos-and-video/video/2011/05/19/open-questions-president-s-speech-middle-east-and-north-africa?page=1	UNCATEGORIZED CLIMATE,HEALTH,SECURITY
https://www.whitehouse.gov/the-press-office/2014/02/11/fact-sheet-us-france-security-cooperation	https://www.whitehouse.gov/the-press-office/2014/02/11/fact-sheet-us-france-security-cooperation	SECURITY UNCATEGORIZED
https://www.whitehouse.gov/omb/legislative_sap_109-1_index-apps	https://www.whitehouse.gov/omb/legislative_sap_109-1_index-apps	UNCATEGORIZED CLIMATE
https://www.whitehouse.gov/legislation/hr-1308-ronald-reagan-centennial-act-amendments-termination	https://www.whitehouse.gov/legislation/hr-1308-ronald-reagan-centennial-act-amendments-termination	UNCATEGORIZED CLIMATE
https://www.whitehouse.gov/omb/2127_meeting_07282011	https://www.whitehouse.gov/omb/2127_meeting_07282011	CLIMATE UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/12/01/president-obama-holds-multilateral-meeting-heads-state?page=8	https://www.whitehouse.gov/photos-and-video/video/2015/12/01/president-obama-holds-multilateral-meeting-heads-state?page=8	UNCATEGORIZED CLIMATE
https://www.whitehouse.gov/photos-and-video/video/2015/12/01/president-obama-holds-multilateral-meeting-heads-state?page=4	https://www.whitehouse.gov/photos-and-video/video/2015/12/01/president-obama-holds-multilateral-meeting-heads-state?page=4	UNCATEGORIZED CLIMATE
https://www.whitehouse.gov/photos-and-video/video/2015/09/22/president-obama-first-lady-vice-president-biden-dr-biden-greet-his	https://www.whitehouse.gov/photos-and-video/video/2015/09/22/president-obama-first-lady-vice-president-biden-dr-biden-greet-his	UNCATEGORIZED CLIMATE
https://www.whitehouse.gov/photos-and-video/video/2015/09/22/president-obama-first-lady-vice-president-biden-dr-biden-greet-his	https://www.whitehouse.gov/photos-and-video/video/2015/09/22/president-obama-first-lady-vice-president-biden-dr-biden-greet-his	UNCATEGORIZED CLIMATE
https://www.whitehouse.gov/blog/2015/02/13/affordable-care-act-enrollment-opportunity-american-indians-and-alaska-natives	https://www.whitehouse.gov/blog/2015/02/13/affordable-care-act-enrollment-opportunity-american-indians-and-alaska-natives	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/omb/114/legislative_sap_date_2016	https://www.whitehouse.gov/omb/114/legislative_sap_date_2016	HEALTH SECURITY,CLIMATE
https://www.whitehouse.gov/photos-and-video/video/climate-change-task-force-hawaii	https://www.whitehouse.gov/photos-and-video/video/climate-change-task-force-hawaii	CLIMATE,HEALTH,SECURITY,SECURITY,HEALTH
https://www.whitehouse.gov/the-press-office/2015/04/17/fact-sheet-va-vet-centers-and-administration-progress-mental-health-0	https://www.whitehouse.gov/the-press-office/2015/04/17/fact-sheet-va-vet-centers-and-administration-progress-mental-health-0	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/blog/2010/01/19/nancy-sutley-focuses-energy-efficiency-new-york	https://www.whitehouse.gov/blog/2010/01/19/nancy-sutley-focuses-energy-efficiency-new-york	CLIMATE UNCATEGORIZED
https://www.whitehouse.gov/blog/2012/03/19/breaking-it-down-health-care-law-and-seniors	https://www.whitehouse.gov/blog/2012/03/19/breaking-it-down-health-care-law-and-seniors	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/the-press-office/2014/11/11/us-china-joint-announcement-climate-change	https://www.whitehouse.gov/the-press-office/2014/11/11/us-china-joint-announcement-climate-change	CLIMATE UNCATEGORIZED
https://www.whitehouse.gov/legislation/streamlining-claims-processing-federal-contractor-employees-act	https://www.whitehouse.gov/legislation/streamlining-claims-processing-federal-contractor-employees-act	CLIMATE
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=18	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=18	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2016/07/22/advancing-21st-century-policing-briefing?page=4	https://www.whitehouse.gov/photos-and-video/video/2016/07/22/advancing-21st-century-policing-briefing?page=4	CLIMATE UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2016/07/22/advancing-21st-century-policing-briefing?page=8	https://www.whitehouse.gov/photos-and-video/video/2016/07/22/advancing-21st-century-policing-briefing?page=8	CLIMATE UNCATEGORIZED
https://www.whitehouse.gov/omb/oira_2066_meetings_191	https://www.whitehouse.gov/omb/oira_2066_meetings_191	CLIMATE UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=10	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=10	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=12	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=12	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=13	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=13	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=16	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=16	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=17	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=17	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=18	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=18	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=19	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=19	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=20	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=20	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=21	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=21	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=22	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=22	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=23	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=23	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=24	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=24	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=25	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=25	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=26	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=26	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=27	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=27	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=28	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=28	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=29	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=29	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=30	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=30	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=31	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=31	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=32	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=32	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=33	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=33	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=34	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=34	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=35	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=35	HEALTH UNCATEGORIZED
https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=36	https://www.whitehouse.gov/photos-and-video/video/2015/06/11/west-wing-week-061215-or-walking-and-holding-hands?page=36	HEALTH UNCATEGORIZED

And sentences!

- We get lists of annotated sentences, in the format:
- SENTENCE, OBAMA-
CLASSIFIED TOPIC, TRUMP-
CLASSIFIED TOPIC



The screenshot shows a GitHub repository page for 'warcbase2017w'. The repository has 4 issues, 1 pull request, 0 projects, and 0 contributors. The file 'diffTrumpPerSentence.txt' contains 542 lines (541 sloc) and is 153 KB. The content of the file is as follows:

```
1 91702 -LRB- Dec. 19 , 2016 -RRB- , promulgated by the Social Security Administration -LRB- SSA
2 President Donald J. Trump Proclaims February as American Heart Month | whitehouse.gov Latest Ne
3 Q You didn t mention climate in your opening statement . (CLIMATE) (UNCATEGORIZED)
4 When you follow a link to an external site , you are leaving WhiteHouse.gov and are subject to
5 Vice President Mike Pence On NBC s Meet The Press : President Trump Has Made It Clear He 's Goi
6 So we ll see where we would reallocate any of that money , but I think the priority is clear i
7 In addition to being good for our economy , boosting domestic energy production is in America s
8 There 's a big difference between Islam , the religion , and radical Islamic terrorists that co
9 There s a lot of individuals -- he s assembled a world-class Cabinet that has a lot of interest
10 PRESIDENT TRUMP : I feel totally confident that we will have tremendous security for the people
11 MR. SPICER : Well , I think the President has made very clear with respect to energy policy that
12 Too much has happened over the past month and in your country and you ; too many new and someti
13 So her admiring new daughter-in-law , Priscilla Cooper Tyler , described her -- `` the most ent
14 There 's reductions , for example , I think , in the NIH -- the National Institutes of Health .
15 This is -- Q This one specifically -- MR. SPICER : Well , again , Cecilia , if I -- I think he
16 President Trump Convened A Meeting With Experts From The Cybersecurity Field To Discuss The Th
17 And North Korea has went on with the missiles and nuclear development . (SECURITY) (UNCATEGORIZ
18 These joint resolutions of disapproval would nullify two rules promulgated by the Employee Bene
19 It is the sense of Congress that advancing propulsion technology would improve the efficiency o
20 Meanwhile , Federal employee health and retirement benefits continue to be based on antiquated
21 So national security is critical in our energy picture . (SECURITY,CLIMATE) (SECURITY)
22 DHS employs 216,000 people in its mission to patrol borders , protect travelers and our transpor
23 Ms. MacDougall was designated acting Chair of the Occupational Safety & Health Review Commission
24 Q There have been some discussions over the last couple of days about whether the U.S. governme
25 Rebuilding Our Nation 's Infrastructure | whitehouse.gov (SECURITY) (UNCATEGORIZED)
26 Dying of ill health during the Revolutionary War , she did not live to see her husband become P
27 The two leaders underscored the importance of shared values between the United States and Brazil
28 For example . the House Committee on Ways and Means includes subcommittees on Social Security a
```

Common Security Sentence

And I think some of the announcements today indicate the fruits of that effort , which is that the President has pushed his national security team to determine if additional **travel restrictions** could be put in place that would make the American public more safe. (SECURITY)

(Top-ranked security sentence)

Common Security Sentence

And there is a `` To-do '' list that the President has put forward to Congress of items that would have a tangible , direct effect on economic growth , a tangible , direct effect on the security -- the **economic security** of the middle class .

(SECURITY)

(Demonstrates flexibility of security concept)

Common Health Sentence

Reflecting on my own experience with cancer , here s what I ve learned : Health care is not just about medicine and treatment , but more importantly about people who are making real-time , human decisions. (HEALTH)

Common Finance Sentence

In fact , the number of people in poverty fell by 3.5 million from 2014 to 2015 , with the poverty rate decreasing faster than at any point since the 1960s -- all racial and ethnic groups saw increases in household incomes and decreases in poverty in 2015 , with African American and Hispanic households seeing the largest declines in poverty . (FINANCE)

Common Climate Sentence

The President committed to lead international efforts to address climate change : Secured the Paris Agreement , where more than 190 countries agreed to a framework for global action on climate change . (CLIMATE)

Different Health Sentence

In March of 2014 , the World Health Organization reported the first outbreak of Ebola in Guinea . (HEALTH)
(UNCATEGORIZED)

Ebola doesn't appear on Trump white house, so classifier could not learn that Ebola = Health (as it did with the Obama White House).

Different Climate Sentence

Man behind the study , professor Jyrki Kauppinen says that climate is indeed warming but only 5 % -10 % of this warming can be addressed to increased greenhouse gas emissions , while IPCC says the warming is almost completely man-made .
(CLIMATE) (UNCATEGORIZED)

Trump doesn't talk much about climate change, so "man-made" never gets picked up by a Trump classifier.

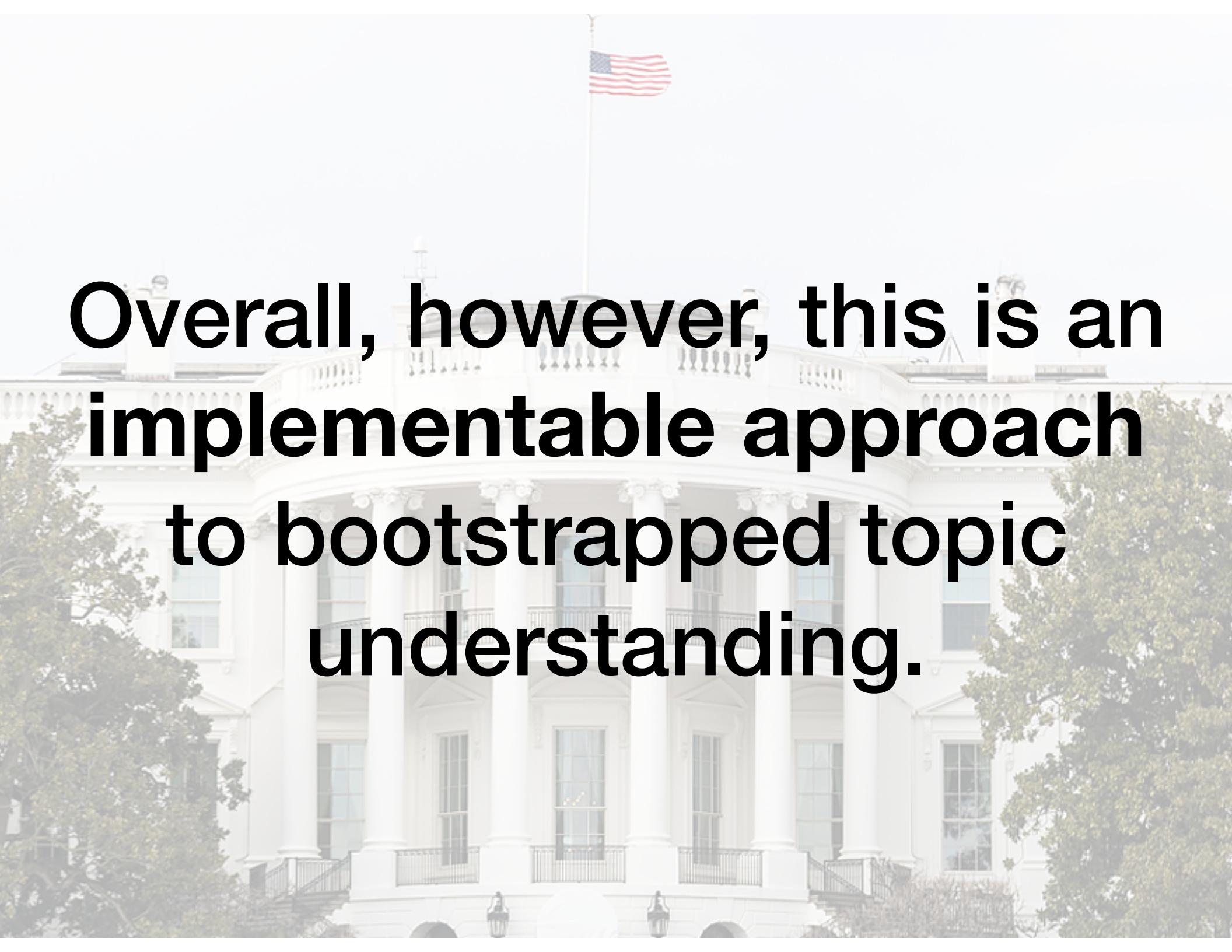
**The machine has learned the
biases of the administrations –
another reminder that
algorithms aren't neutral.**

Advantages for humanists

- Why do I like this?
 - Topic modelling, for example, is hard to sell to peer reviewers – it's got a bit of a black box for a humanist or social scientist.
 - **I can explain this method and understand it – it melds the “bag of words” approach that digital humanists are familiar with to machine learning algorithms**
- There is no black box, at least to somebody like me.

Disadvantages for humanists

- There are of course some disadvantages:
 - Need a “**bag of words**,” either from a subject-matter expert or from your own research team;
 - The **biases of the classifier** are put into it - a Trump-trained classifier is different than Obama;



**Overall, however, this is an
implementable approach
to bootstrapped topic
understanding.**



**Bootstrapped topic detection
in an easy-to-understand and
cheap workflow.**



**NSERC
CRSNG**



Social Sciences and Humanities
Research Council of Canada

Conseil de recherches en
sciences humaines du Canada



UNIVERSITY OF
WATERLOO

Canada



Thanks!

A large, bold, black sans-serif font "Thanks!" is centered over a background image of a large, white, classical-style building with multiple columns and a balcony. The building is partially obscured by green trees on either side.

**Ziquan Wang
Borui Lin
Ian Milligan
Jimmy Lin**

