

A smiling man with dark hair, wearing a dark jacket over a light-colored shirt, stands in front of a wooden trellis structure. The trellis is covered with climbing plants that have small purple flowers. The background is slightly blurred, showing more of the trellis and foliage.

How One Metis Student Made a Tabloid News Classifier.

Topic modeling clickbait in a world of information overload.

Andrew Tom | June 2, 2017 | [awww slack!](http://awww.slack!) [@andrewtom](https://twitter.com/andrewtom)

“A fiercely independent press remains fundamental to the stability of our society...Journalists represent the people as much as electeds.”



Dan Rather, former host of 60 minutes.

Media Consolidation: THE ILLUSION OF CHOICE

Media has never been more consolidated. 6 media giants now control a staggering **90%** of what we read, watch, or listen to.

THESE SIX COMPANIES ARE:



GE

Notable
Properties:

COMCAST
NBC
UNIVERSAL
PICTURES
FOCUS
FEATURES



**NEWS-
CORP**

Notable
Properties:

FOX
WALL STREET
JOURNAL
NEW YORK
POST



DISNEY

Notable
Properties:

ABC
ESPN
PIXAR
MIRAMAX
MARVEL
STUDIOS



VIACOM

Notable
Properties:

MTV
NICK JR
BET
CMT
PARAMOUNT
PICTURES



**TIME
WARNER**

Notable
Properties:

CNN
HBO
TIME
WARNER
BROS



CBS

Notable
Properties:

SHOWTIME
SMITHSONIAN
CHANNEL
NFL.COM
JEOPARDY
60 MINUTES

Exclusive To All Newspapers

DIANA WAS ALIVE HOURS BEFORE SHE

A SERIES of previously unseen CCTV images have revealed that Diana was alive and well before she was tragically killed in Paris, ten years ago.

INSIDE • Do you think Diana is dead? Or have you seen her in Morocco? Call now: 0900 000 0000. If not, call: 0900 000 0000. Calls cost (cont. p. 94)

BREAKING TRUMP TWEET

res MSCI Mexico (EWW) 45.35 ▼ 0.11 Whirlpool (WHR) 179.94 ▼ 10.26
Comcast (CMCSA) 73.60 ▲ 0.16 Mattel (MAT) 27.15 ▼ 4.41 Norwegiar

THE INTELLIGENCE REPORT

WHITE HOUSE: WE HAVE A POROUS BORDER WITH BROKEN FENCES

MARKET ALERT

5:56A PACIFIC

THE WHITE HOUSE
2:56pm ET

DOW
20,953.29
▲ 3.40 0.02%

MARKET ALERT

WHITE HOUSE: WE HAVE A POROUS BORDER WITH BROKEN FENCES

NIKKEI 19,402.39
SHANGHAI 3,159.17
HANG SENG 23,374.17

SILVER 16.58 ▼ 0.25 S&P 2386.98 ▼ 4.19



Journalists are fallible.

Problem:

Illustrate distinctions between

broadsheet and tabloid journalism.

What Can We Measure? The Spectrum of Print Journalism

Broadsheets

The New York Times

Friday, June 2, 2017 11:00 AM EDT Today's Paper Video FTSE 100 +0.31% 1

World U.S. Politics N.Y. Business Opinion Tech Science Health Sports Arts Style Food Travel Magazine T Magazine Real Estate All

Trump Pulls U.S. Out of Climate Accord, Saying It's a Threat to U.S. Economy

Environment Will Suffer as Result, Allies and Rivals Say

By MICHAEL D. SHEAR

- In choosing to leave the pact, President Trump sided with G.O.P. conservatives. "I was elected to represent the citizens of Pittsburgh, not Paris."
- The decision was a rebuke to heads of state, climate activists, corporate executives and even some members of the president's own staff.

4429 Comments

International Call to Action

ENVIRONMENTAL NEWS

Void in Leadership Opens Door to Global Upheaval

By DAVID E. BARNER and JAMES PEARCE

The decision is perhaps the greatest strategic gift to the Chinese, who are eager to expand their diplomatic reach and finance the infrastructure projects that give Beijing ever-greater influence.



NEWS ANALYSIS

A 'Draconian' Pact? The Facts Say Otherwise

By MARK LANDLER, BRAD PLUMER and LINDA QIU 10:54 PM ET

The president seemed to base his decision more in his belief in its potential to save jobs than in any disbelief in climate science. But he also justified his action with distorted reports.

Cities, States and Businesses Circumvent the President

By ROBERT FARBUSCH and HENRY ROYDEN

A group of American officials, intellectuals and businesses is negotiating with the U.N. to accept its emissions plan.

• Corporations Push Back: 'Not Good for America'

NEWS ANALYSIS

Looking to 2018 (and Beyond), Trump Bets on His Base

By PETER BAKER

He gambled that placating his most ardent fans would serve him

The Opinion Pages

EDITORIAL

Our Disgraceful Exit

Here's what Mr. Trump's decision on the climate-change pact says to the world: America cares little about science, its allies and competitiveness.

Trump's Stupid and Reckless Climate Decision

By STEVE BERNARD

Quitting the climate accord undermines the planet's best hope.

• Brooks: Trump Poisons the World

• Krugman: Trump Graciously Rejects the Paris Accord

• The World Is No Protection From Toxic Chemicals

TWICE INSIDER

Inside 'The Daily' With Michael Barbaro

THE 50 DAY WAR AT ISRAELI OCCUPATION. AND THE NEXT.

By NATHAN TREBALL

Israel enjoys perpetual control of the Palestinian land it captured in 1947; thanks to myth and military power.

The 50 Day War at Israeli Occupation. And the Next.

By NATHAN TREBALL

Israel enjoys perpetual control of the Palestinian land it captured in 1947; thanks to myth and military power.

• On Campus When the Left Turns on Its Own

• The Scope of Hate in 2017

THE CROSSWORD

Play Today's Puzzle

3H George Soros accused Hungary's prime minister, Viktor Orban, of running a "mafia state." Mr. Orban threatened to shutter a university that Mr. Soros founded in Budapest.



5H "It feels like history, of course." At the French Open, one Jabeur became the first woman from an Arab country to reach the third round of a major tennis tournament.



Tabloids

ENQUIRER

MEGYN KELLY

O.J. SIMPSON

SEARCH

FACEBOOK

TWITTER

GOT A TIP?

GOTCHA!

BILL COSBY BLINDNESS SCAM — BUSTED BY NATIONAL ENQUIRER

STAKEOUT SHOTS DOWN PERVERT'S CLAIMS AS SEX ATTACK TRIAL LOOMS!



COURT RULING!

JIM CARREY TO FACE TRIAL OVER EX-GIRLFRIEND'S DEATH

Star still haunted by tragic suicide!



FAMILY PHOTOS!

BRAD PITT & SON ATTEND FAMILY THERAPY

Superstar struggling to stay involved!

GET OUR EXCLUSIVE OFFER!



SUBSCRIBE NOW

SUBSCRIBE

GIVE A GIFT

RENEW

CUSTOMER SERVICE

What Can We Measure? The Spectrum of Print Journalism

Broadsheets

- Disseminate news to the public
- Traditional newsgathering and longer articles
- Focus on 'serious' topics like politics and international news

Tabloids

- Deliver interesting content
- Tend to more racy and sensationalistic
- Focus on 'less serious' topics like culture and celebrity gossip

Unleash the NLP:

**Confirm the distinctions between
broadsheet and tabloid journalism via
clustering and topic modeling.**

Process

Step 1

Scraped 80,000+ articles
(broadsheet and tabloid)
→ TFIDF Vect



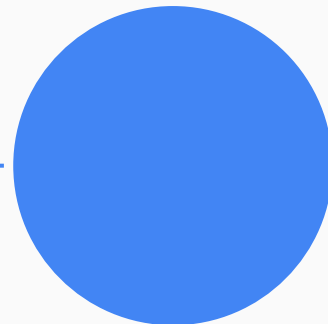
Step 2

Multinomial Naive Bayes
Score: 93%
Language is distinct



Step 3

HDP → LDA over each corpus
(20 Topics, 3 Passes):
Topics more descriptive



Broadsheet Corpora
(2005–2015)



Latent topics:

- Economy
- Context
- Politics

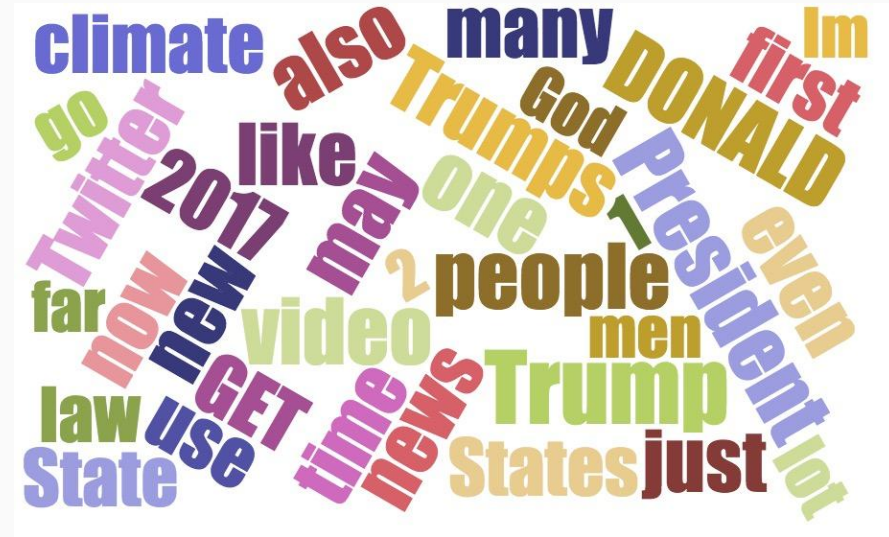
Top 3 Latent Topics for the Tabloid Corpora (2017)

“Whenever trump communicates, things get sensational.”

Latent Topic 2: (21.8%)	0.012*"heat" + 0.012*"feed" + 0.012*"delivered" + 0.011*"reading" + 0.011*"street" + 0.010*"latest" + 0.010*"thanks" + 0.009*"like" + 0.008*"trump" + 0.006*"video"
Latent Topic 13: (10.0%)	0.008*"trump" + 0.003*"president" + 0.003*"said" + 0.003*"news" + 0.003*"information" + 0.002*"video" + 0.002*"people" + 0.002*"great" + 0.002*"right" + 0.002*"donald"
Latent Topic 5: (8.2%)	0.013*"trump" + 0.005*"president" + 0.003*"reading" + 0.003*"did" + 0.003*"feed" + 0.003*"heat" + 0.003*"delivered" + 0.003*"spicer" + 0.003*"twitter" + 0.003*"street"

Latent topics:

- Anytime Trump communicates



Tabloid Corpora
(2017)

Top 3 Latent Topics for the Broadsheet Corpora (2005-2015)

Latent Topic 8: Economy (13.5%)	0.012*"percent" + 0.011*"year" + 0.009*"said" + 0.008*"billion" + 0.007*"company" + 0.007*"sales" + 0.006*"quarter" + 0.006*"million" + 0.005*"bank" + 0.005*"market"
Latent Topic 15: Issues in Context (7.0%)	0.007*"new" + 0.005*"year" + 0.004*"like" + 0.004*"company" + 0.003*"city" + 0.003*"school" + 0.003*"old" + 0.003*"prime" + 0.003*"san" + 0.003*"people"
Latent Topic 13: Politics (9.1%)	0.015*"president" + 0.012*"obama" + 0.011*"house" + 0.009*"state" + 0.009*"democrats" + 0.008*"republican" + 0.008*"senate" + 0.007*"republicans" + 0.007*"health" + 0.007*"tax"

law
Obama
two
like ago
also
last
States
company
One
percent
first
new
cut
bA
Year
years
far
tax
President
The
big
oil
day
part
policy

Tabloid Corpora
(2017)

Two word clouds demonstrating word frequency across the broadsheet and tabloid corpora.

Future Work

Show where you are in the process and what's left to tackle





DECEMBER 12, 2013 • VOL. 49 ISSUE 50

AMERICA'S FINEST NEWS SOURCE • ONION.COM

Copyright © 2013 Onion, Inc. All Rights Reserved.

INSIDE

INK *The Rise Of The Most Enduring And Trusted Player In Modern Journalism*

PART ONE OF AN EIGHT-PART SERIES

HIGHLIGHTS



ation Just Prefers Feel
Newsprint In Hands
TIONAL, Page 148



'ONION' PRINT REVENUES UP 5,000%



Demand for digital content "simply isn't there yet," experts say.

Experts: Digital Media Revolution Still Another 70 Or 80 Years Away

NEW YORK—Confirming that the current media landscape has so far shown virtually no sign of moving away from the lucrative print trade, a group of experts told reporters today that the digital revolution is still another 70 or 80 years away.

INSIDE

INK *The Rise Of The Most Enduring And Trusted Player In Modern Journalism*

PART ONE OF AN EIGHT-PART SERIES

What we care about is what we'll get.

'ONION' PRINT REVENUES UP 5,000%



Demand for digital content "simply isn't there yet," experts say.

Experts: Digital Media Revolution Still Another 70 Or 80 Years Away

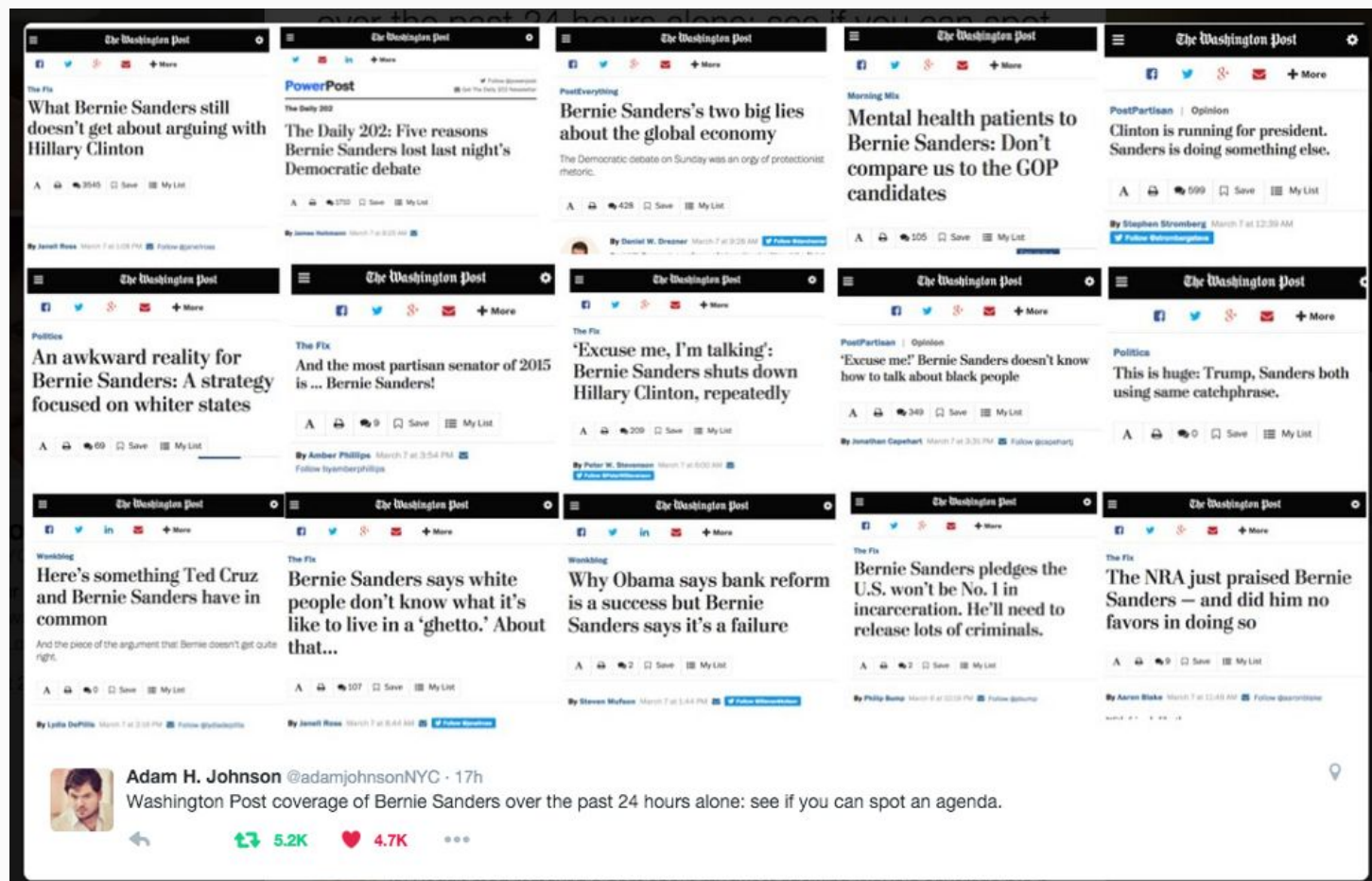
NEW YORK—Confirming that the current media landscape has so far shown virtually no sign of moving away from the lucrative print trade, a group of experts told reporters today that the print industry is still another 70 or 80 years away from being replaced by digital content.

Questions?

Appendix

Assumptions & Limitations

- That every article from a given source adheres to the ‘seriousness’ or ‘sensationalist’ genre of the publication.
 - Clearly not true
 - The corpora covered different time intervals
 - Topic modeling was not comparable
 - No intuition for sensitivity along the ‘print media’ spectrum
-



A collection of clickbait-y headlines from the New York Times in 2016. (P.S. I Scraped from 2005-2015)

2009: Michael Jackson had "six months to live."

In January 2009, the Enquirer **reported that** pop superstar Michael Jackson had only six months to live because of drug and alcohol addiction.

The singer **passed away six months later**, in June.



Source Selection

Broadsheet Sources

- New York Times
- Reuters (via circular reporting)
- Associated Press (via circular reporting)

Tabloids

- 30+ Sources including:
 - National Enq
 - addictinginfo.org
 - liberalamerica.org
 - politicususa.com
 - viraltube.nl
 - And more from Melissa Zimdars' list.

Melissa Zimdars is an assistant professor of communication and media at Merrimack College.

The Real Process

I can't even word cloud

Yo this HDP is 🔥

LDA FTW

Step 1

Multinomial Naive Bayes
Score: 93%

Step 1.7: I hate AWS

Step 2

Hierarchical Dirichlet Process
over each corpus (150 topics):
Topics not descriptive

Step 3

LDA over each corpus
(20 Topics, 3 Passes):
Topics more descriptive

FINALLY THIS
MAKES SENSE

OMG MemoryErrors

Did an LSI, or 2

To aggregate or not
to aggregate?

Wtf is NLP

