Getting to the Core of Algorithmic News Curators: A Case Study of Apple News

Jack Bandy (Northwestern University) @jackbandy
Parts of Siberia, Mars and the South Pole all will be warmer than Chicago this week.
My Research Ideation Machine (Beta)

Is (platform or technology) making us (vice or trait)?

Facebook
Google Search
Netflix
Google Ads
GPS
YouTube
Reddit
Apple News

antisocial
intolerant
biased
impatient
naive
uninformed
Apple News

TRENDING STORIES

1. CNN politics
   Here's why House Democrats are going to go after Trump's taxes -- and soon
   4h ago

2. FOX NEWS
   Wisconsin woman found frozen under vehicle in hospital parking garage dies, suspect arrested
   8h ago

3. People
   Camille Grammer Is Living in a 'Beautiful Trailer-Home' After Her Mansion Burned Down in Wildfire
   4h ago

4. BUSINESS INSIDER
   Russian media says Su-35 fighter jets 'humiliated' and backed down US F/A-18s
   3h ago

TOP STORIES

1. NEWS
   Trump accepts Pelosi invitation to deliver State of the Union on Feb. 5
   48m ago

2. FOX NEWS
   5 police officers injured in Houston shooting, officials say
   8h ago

3. TRENDING
   CNN politics
   Here's why House Democrats are going to go after Trump's taxes -- and soon
   4h ago

4. FOX NEWS
   Wisconsin woman found frozen under vehicle in hospital parking garage dies, suspect arrested
   48m ago

TOP VIDEOS

1. BBC NEWS
   Kite surfer 'jumps' 200m over sand spit
   1 OF 6 VIDEOS
Related Work

• Apple News
  • Columbia Journalism Review (Brown)
  • New York Times (Nicas)

• Algorithm Audits (Sandvig)
  • Underrepresentation (Kay)
  • Filter Bubbles (Bakshy)

• News Platform Audits
  • Google News (Haim; Nechushtai)
Research Questions

• How do news curation systems like Apple News influence the public’s media intake?

• What is the system’s mechanism?
  • How often does it update?
  • Does it localize or personalize?

• What content does it direct attention to?
  • What sources does it feature?
  • What topics does it feature?
Black Box Dilemma

• Proprietary code
• No public APIs or endpoints
• SSL Pinning
• Possible data collection methods:
  • Apple News Twitter (Brown)
  • Email Newsletters (Brown)
  • Crowdsourcing

```bash
$ ./scrape_apple_news
ERROR
$ ./scrape_apple_news
ERROR
$ ./scrape_apple_news
ERROR
```
Methods: The Crowd

• Amazon Mechanical Turk

• **Pros**
  • Circumvents black box
  • Real-world data
  • High parallelism/throughput

• **Cons**
  • Data Verification
  • Inconsistent coverage
Methods: Appium

• Automated App Control

• **Pros**
  • Lower cost
  • Sustained coverage
  • No manual inspection

• **Cons**
  • Single channel
  • Data points *in vitro*
Findings: Source Concentration

Relative Distribution of Trending Stories
Combined: December 12th-20th, 2018; January 4th-12th, 2019

Fox News  CNN  People  HuffPost  Politico  Newsweek  BuzzFeed  Vanity Fair  Vox  Washington Post

% of Trending Stories (n=576)
## Findings: The Human Touch

<table>
<thead>
<tr>
<th></th>
<th>Trending Stories</th>
<th>Top Stories</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Curation</strong></td>
<td><strong>Algorithmic</strong></td>
<td><strong>Editorial Staff</strong></td>
</tr>
<tr>
<td>Headlines Displayed</td>
<td>4 (6 on big screens)</td>
<td>5</td>
</tr>
<tr>
<td>Localization</td>
<td>National</td>
<td>National</td>
</tr>
<tr>
<td>Personalization</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Avg. Stories / Day</td>
<td>31</td>
<td>16</td>
</tr>
<tr>
<td>Total Stories</td>
<td>279</td>
<td>144</td>
</tr>
<tr>
<td>Total Sources</td>
<td>28</td>
<td>40</td>
</tr>
<tr>
<td>Avg. Stories / Source</td>
<td>9.9</td>
<td>3.6</td>
</tr>
<tr>
<td>Stdev. Stories / Source</td>
<td>14.6</td>
<td>3.3</td>
</tr>
<tr>
<td>#1 Source %</td>
<td>20.1% (Fox)</td>
<td>9.0% (WaPo)</td>
</tr>
<tr>
<td>#1-#3 Sources %</td>
<td>50.5%</td>
<td>25.7%</td>
</tr>
<tr>
<td>#1-#10 Sources %</td>
<td>85.7%</td>
<td>55.6%</td>
</tr>
</tbody>
</table>

Data collected January 4th-12th, 2019
Conclusions & Next Steps

• How does Apple News affect Local and Regional news outlets?

• How do people actually use the app? Do they prefer one section?

• Do similar patterns (source concentration, the human touch) show up in other aggregators?

• Have ideas? Reach out!  @jackbandy  jackbandy.com


