Improving Estimates of Arrest Related Deaths with Machine Learning

October 25, 2018

Government Advances in Statistical Programming (GASP!) Workshop

Peter Baumgartner (Data Scientist @ RTI International)

This project was conducted with funds awarded by the Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. The views, opinions, and content expressed in this exhibition do not necessarily reflect the views, opinions, or policies of the Bureau of Justice Statistics (BJS), the Office of Justice Programs (OJP), or the U.S. Department of Justice (DOJ).

History

2000: Death in Custody Reporting Act (DICRA) passed

2003-2009: Bureau of Justice Statistics (BJS) collects Arrest Related Death (ARD) statistics.

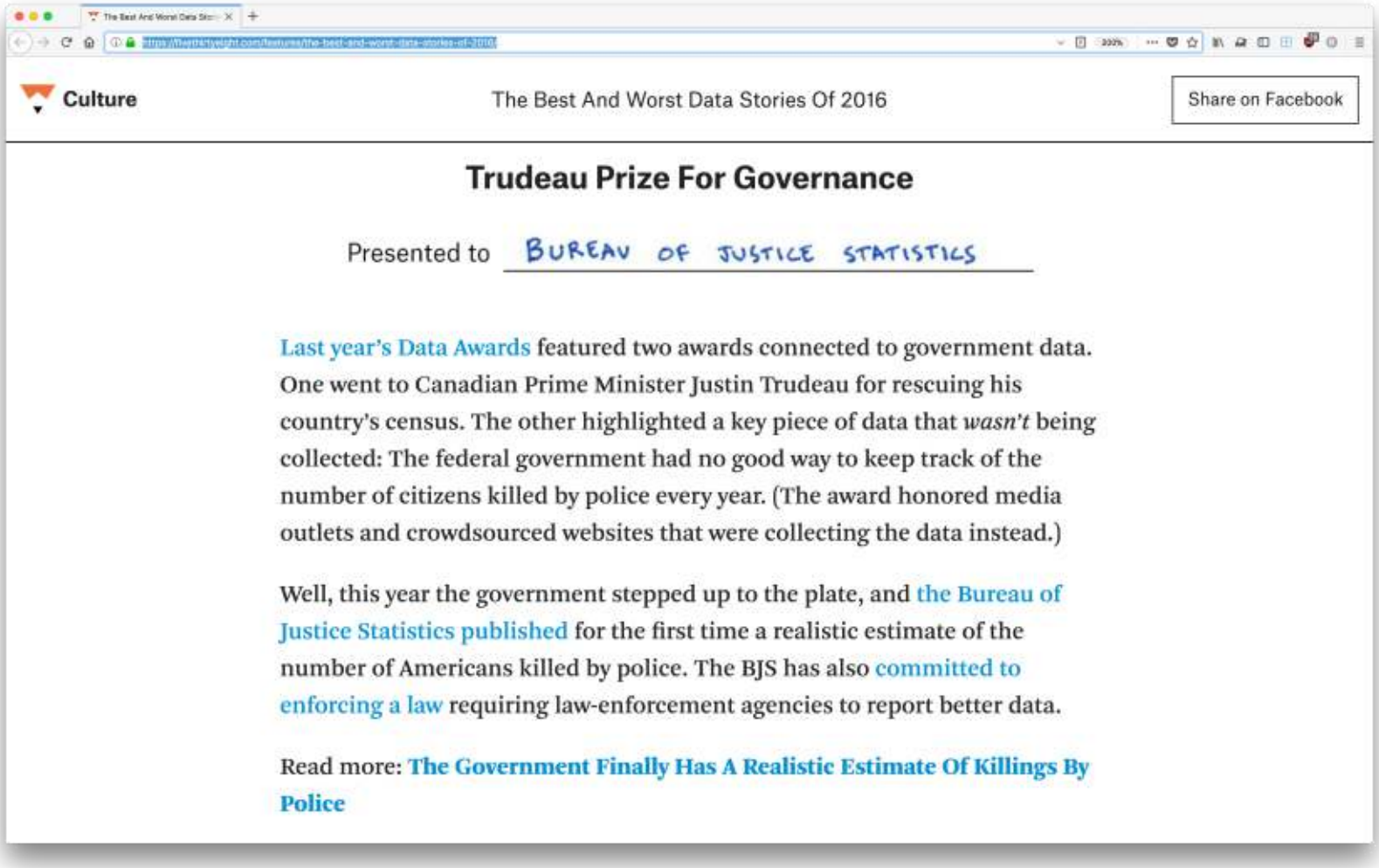
2011: BJS publishes ARD statistics for 2003-2009

2014: BJS conducts a coverage assessment of the ARD program. Results indicate ARD data collection is not a census of relevant incidents.

2015: BJS begins a pilot study to test a revised data collection approach that includes multiple methods of identifying and confirming arrest-related deaths.

https://www.bjs.gov/index.cfm?tid=82&ty=tp

2016



Culture

The Best And Worst Data Stories Of 2016

Share on Facebook

Trudeau Prize For Governance

Presented to BUREAU OF JUSTICE STATISTICS

Last year's Data Awards featured two awards connected to government data. One went to Canadian Prime Minister Justin Trudeau for rescuing his country's census. The other highlighted a key piece of data that *wasn't* being collected: The federal government had no good way to keep track of the number of citizens killed by police every year. (The award honored media outlets and crowdsourced websites that were collecting the data instead.)

Well, this year the government stepped up to the plate, and the Bureau of Justice Statistics published for the first time a realistic estimate of the number of Americans killed by police. The BJS has also committed to enforcing a law requiring law-enforcement agencies to report better data.

Read more: **The Government Finally Has A Realistic Estimate Of Killings By Police**

https://fivethirtyeight.com/features/the-best-and-worst-data-stories-of-2016/

Hybrid Approach to Data Collection

Phase 1: Identification

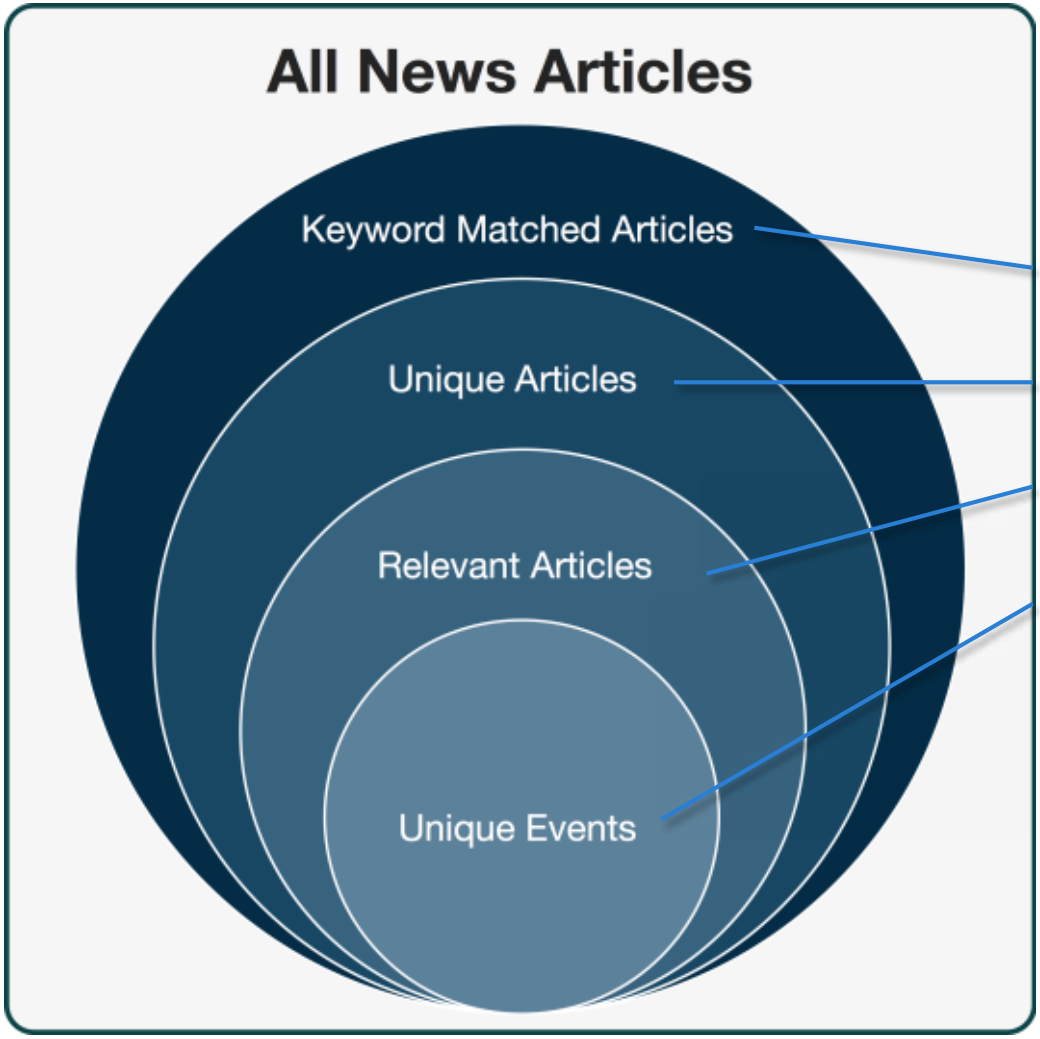
Focus of this talk

- Review news alerts and other public sources for potential arrest-related deaths.
- Track the efficiency and coverage of various approaches for identifying arrest-related deaths.
- Compile quarterly lists of potential arrest-related deaths.

Phase 2: Agency survey

- Conduct follow up with law enforcement agencies and medical examiner/coroner's offices to confirm identified deaths and collect more information about the circumstances surrounding the deaths.
- Survey agencies with identified deaths and others to assess and adjust Phase 1 identification methods.

The Process



A typical month…

1,200,000 articles

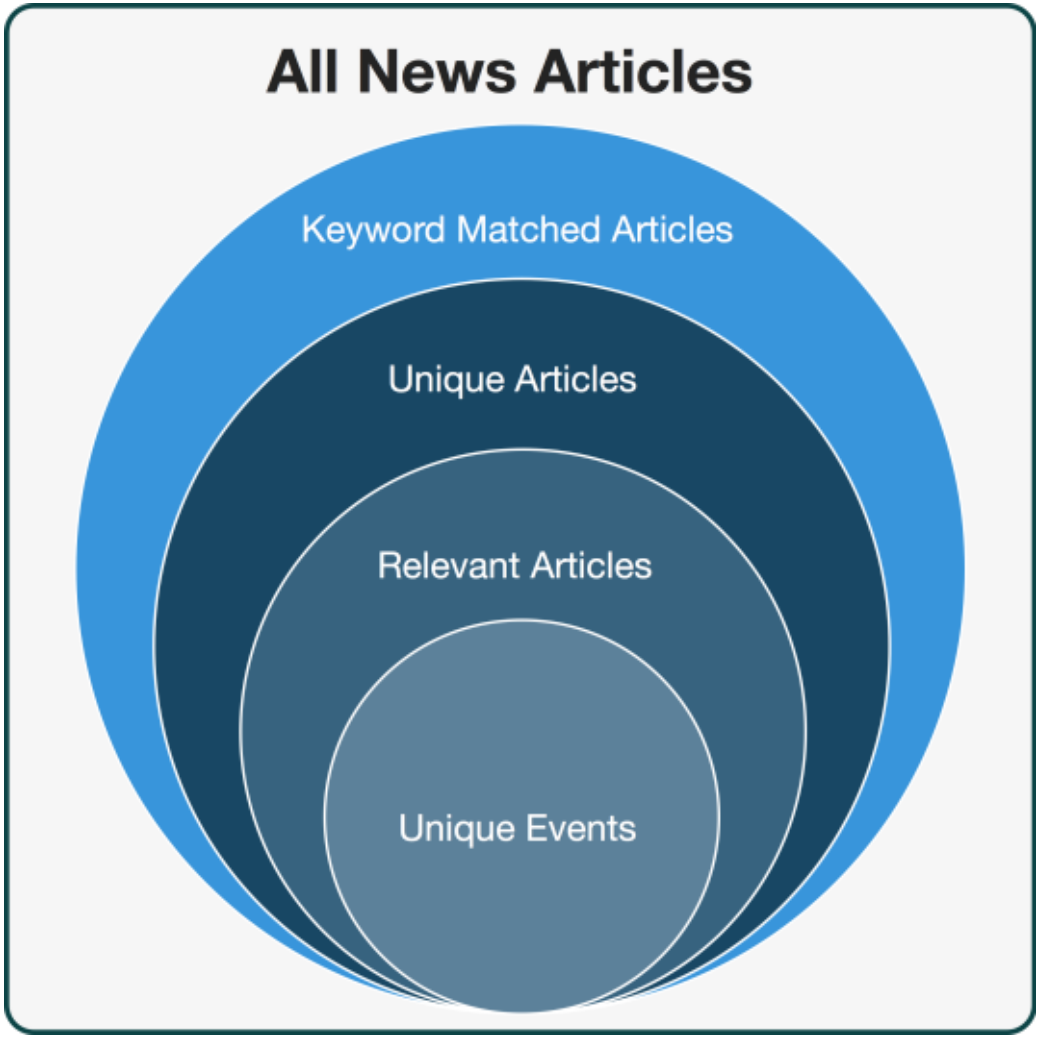
245,000 articles

8,750 articles

135 events

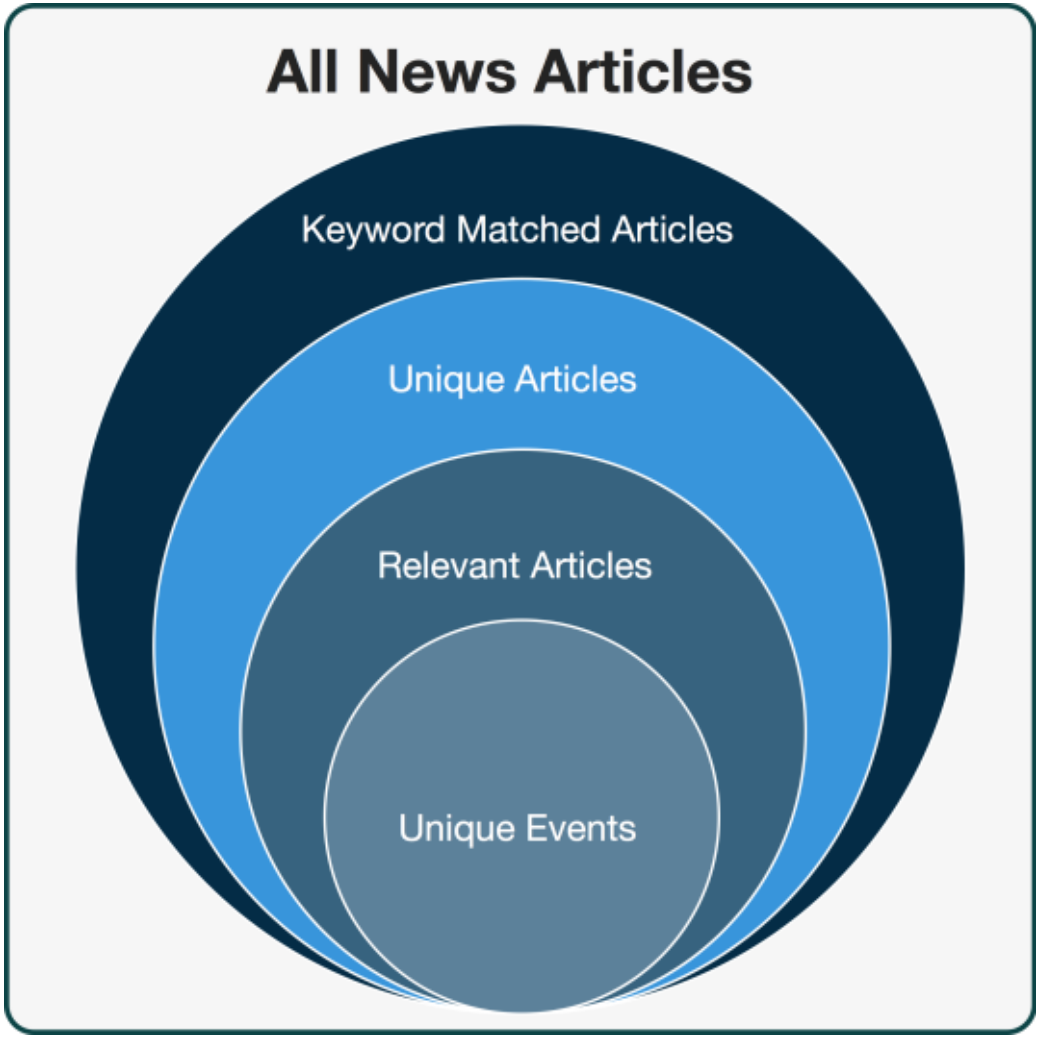
Article Reduction:
99.2%

Keyword Matched Articles



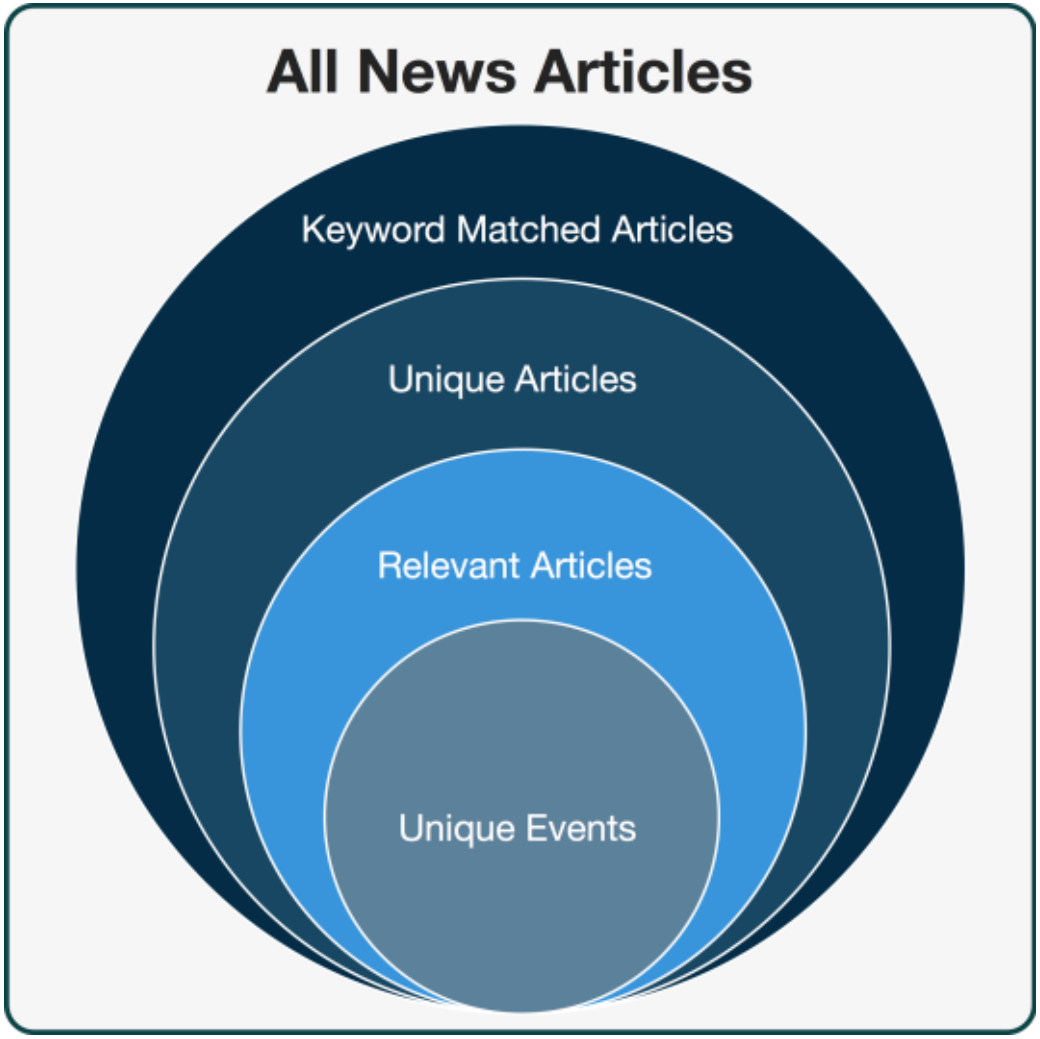
- Keywords are a noun and verb combination indicating law enforcement and relevant action.
- e.g. “officer” & “lethal force”
- Use word embedding models (word2vec) to expand keywords.

Unique Articles



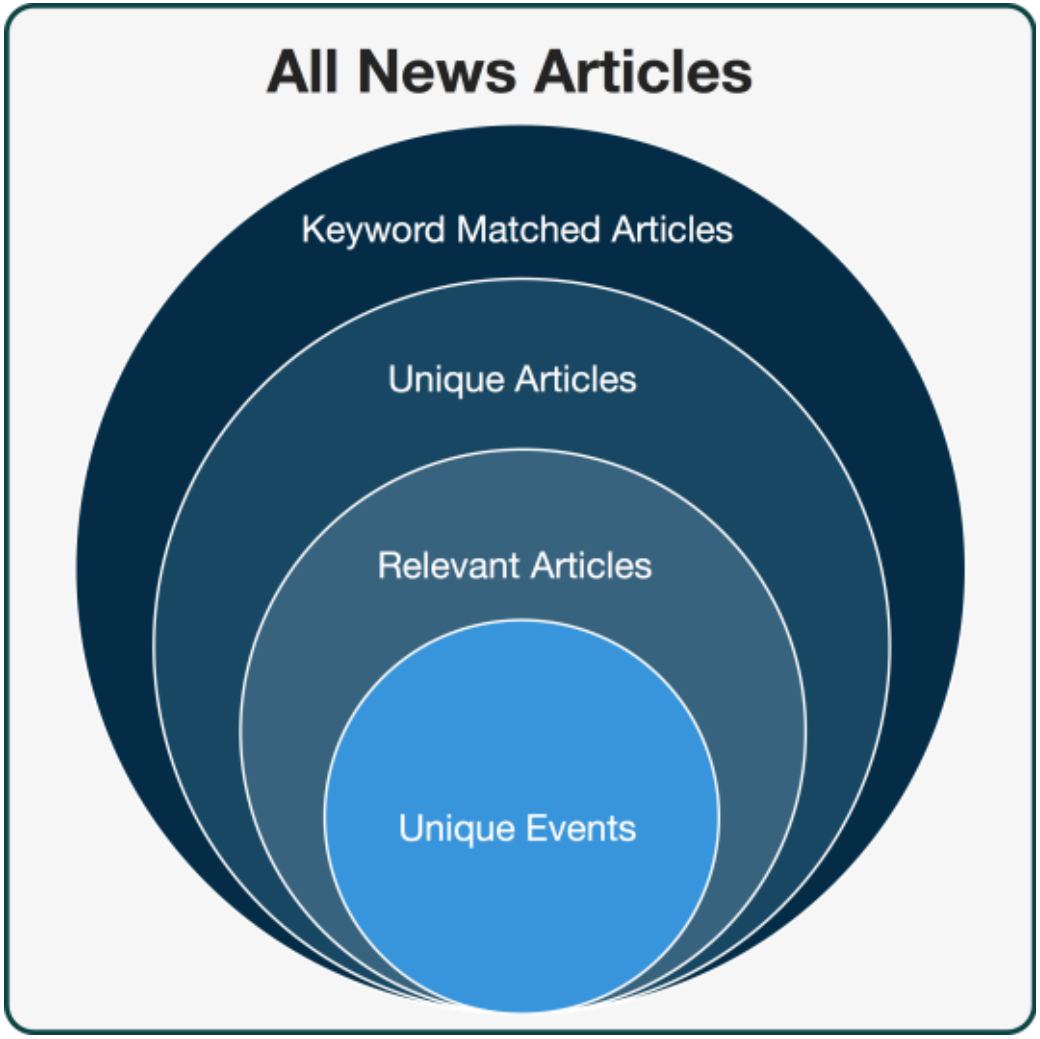
- Remove exact duplicates
- Remove certain sources
- Deduplicate based on Title similarity
- Deduplicate based on Text similarity
- Deduplicate based on Entity similarity

Relevant Articles



- Train article classification model based on labeled articles
- Update article classification model as coders label
- Evaluate model on randomly sampled articles

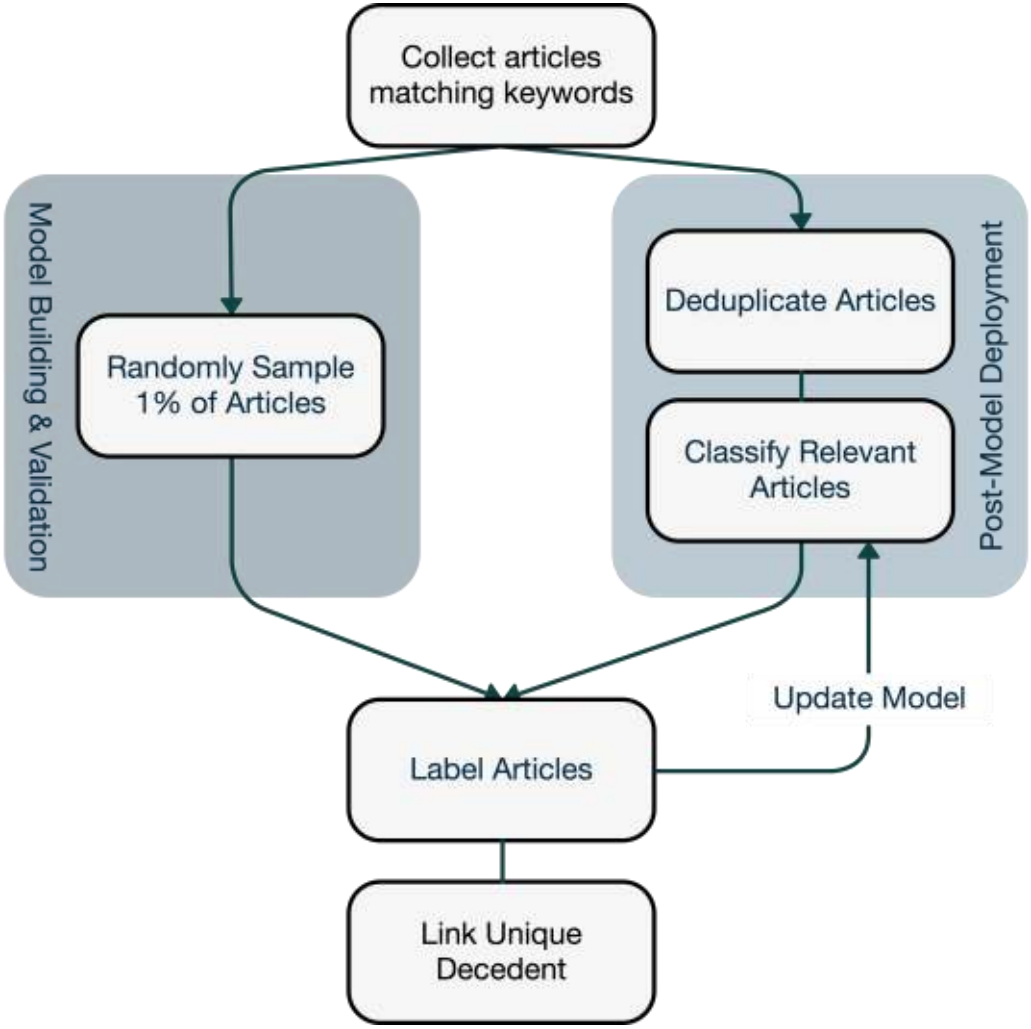
Unique Events



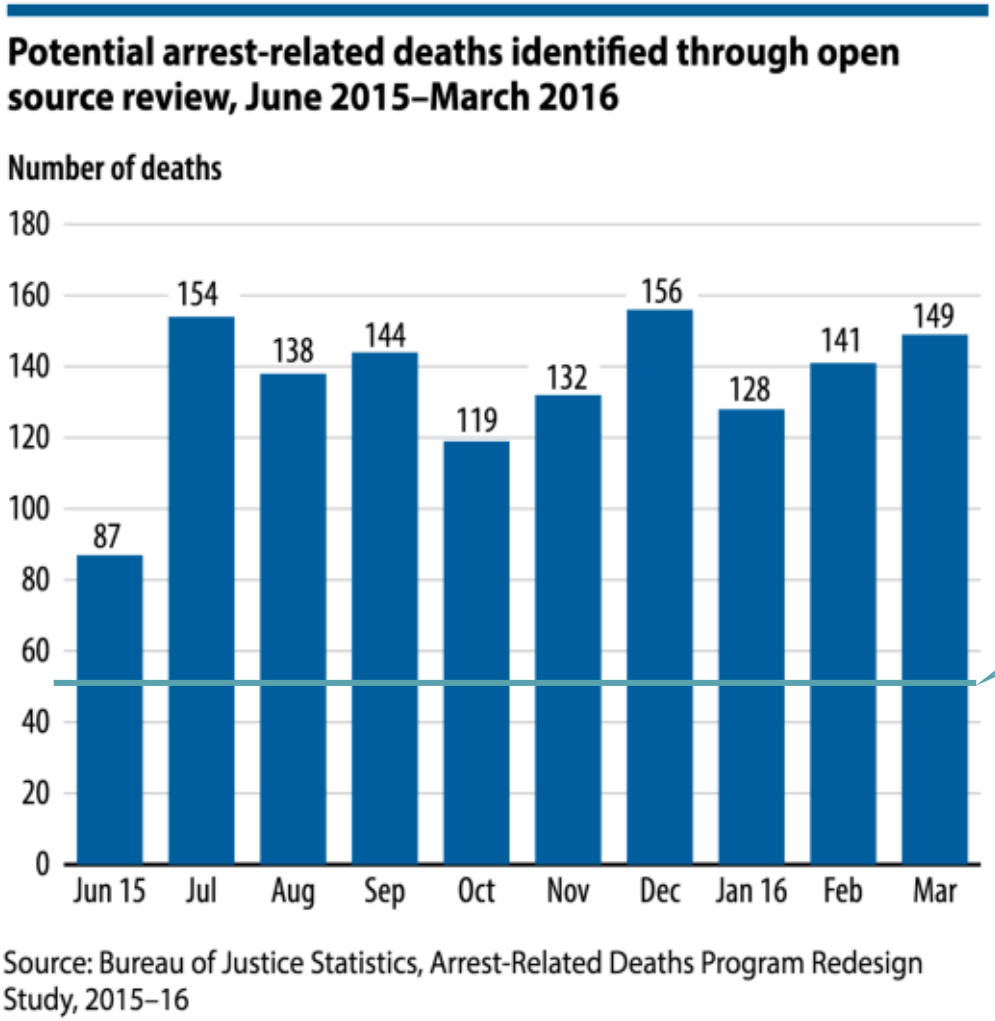
With a web interface:

- A team of coders labels articles meeting the definition of ARD
- Articles about the same event are linked
- Coders enter details about incidents
- QC identifies data issues

Model Updating



Results



Source: Bureau of Justice Statistics, Arrest-Related Deaths Program Redesign Study, 2015–16

Average number of arrest-related deaths identified per month in 2011 through the previous ARD program methodology

https://www.bjs.gov/content/pub/pdf/ardprs1516pf.pdf

Thanks to…

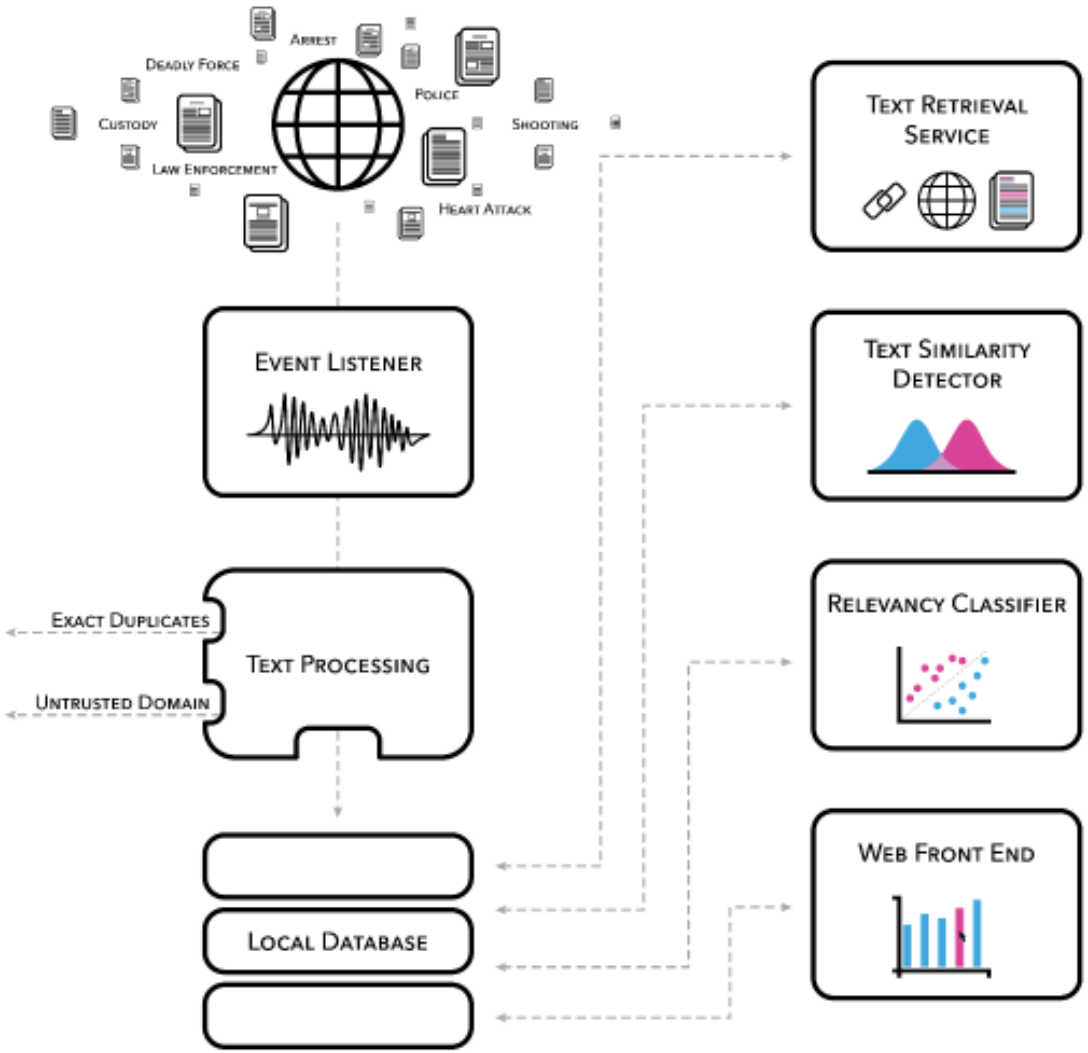
Bureau of Justice Statistics

Duren Banks
Erin Kennedy
Jason Nance
Paul Ruddle

Thank you

Appendix

Behind the scenes…



python™

pandas

scikit learn

Keras

spaCy

gensim
topic modelling for humans

jupyter

PostgreSQL

Text Similarity Detector

A former **South Carolina** police chief has dodged prison time in the 2011 shooting death of an unarmed Black man, receiving just one year of home detention, according to NPR.

Former Eutawville, S.C. police chief **Richard Combs**, 38, shot and killed 54-year-old **Bernard Bailey** as he left the Eutawville Police Department in May 2011.

Source: *NewsOne*

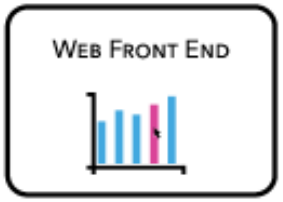
A former South Carolina police chief who was charged with murder in the shooting death of a black man in 2011 pleaded guilty to misconduct charges after two separate trials ended in hung juries.

The former Eutawville chief, Richard Combs, had been indicted more than three years after he shot and killed Bernard Bailey, 53, outside the police department. Combs argued that he feared Bailey would use a vehicle as a weapon against him.

Source: *Washington Post*

1,200,000 → 515,000 → 245,000 articles/month

Web Front End



ARD

Last Decedent Reviewed: [illegible] | Articles Reviewed Today: 3

Duren | Log Out

Agent: Officer had conflicting accounts of black man's death

MONTGOMERY, Ala. — A white Alabama police officer charged with killing a black man gave two conflicting accounts of what happened to a state investigator, the agent testified at a hearing Thursday. Officer Aaron Smith is charged with murder in the ...

Open source article in a new window

- Does Not Meet ARD Definition
- Unknown Whether Meets ARD Definition (Needs Follow-Up)
- Meets ARD Definition
- Meets ARD Definition But Does Not Fall Within Our Date Scope

Save and close

Follow-Up Notes

Notes

First Name

First name

Middle Name

Middle name

Last Name

Last name

Agency Name

Agency name

News | Local and State | Education | Business | Arts | Health | Entertainment | Lifestyle Leisure

24 March 2016 Thursday 17:49 — This news clip was read 29 times.

Agent: Officer had conflicting accounts of black man's death

State Bureau of Investigation Agent Jason DiNunzio testified Smith didn't suspect Gunn of a crime when he initially stopped him, but Smith immediately confronted Gunn and told him to put his hands on the hood of the patrol car. DiNunzio said he doesn't...

Share on Facebook | Tweet | Share on Google+

T- T+ A^

Agent: Officer had conflicting accounts of black man’s death

MONTGOMERY, Ala. — A white Alabama police officer charged with killing a black man gave two conflicting accounts of what happened to a state investigator, the agent testified at a hearing Thursday.

Officer Aaron Smith is charged with murder in the Feb. 25 death of 58-year-old Greg Gunn. After Thursday's hearing, Judge Jimmy Pool ruled there was probable cause

★ Featu

[illegible]egory Gunn | Montgomery, AL | Montgomery PD | 2016-02-25 | (96%)

Select Known Decedent