News Pulse

Data Pipeline, NLP Algorithm & Web App

Goals

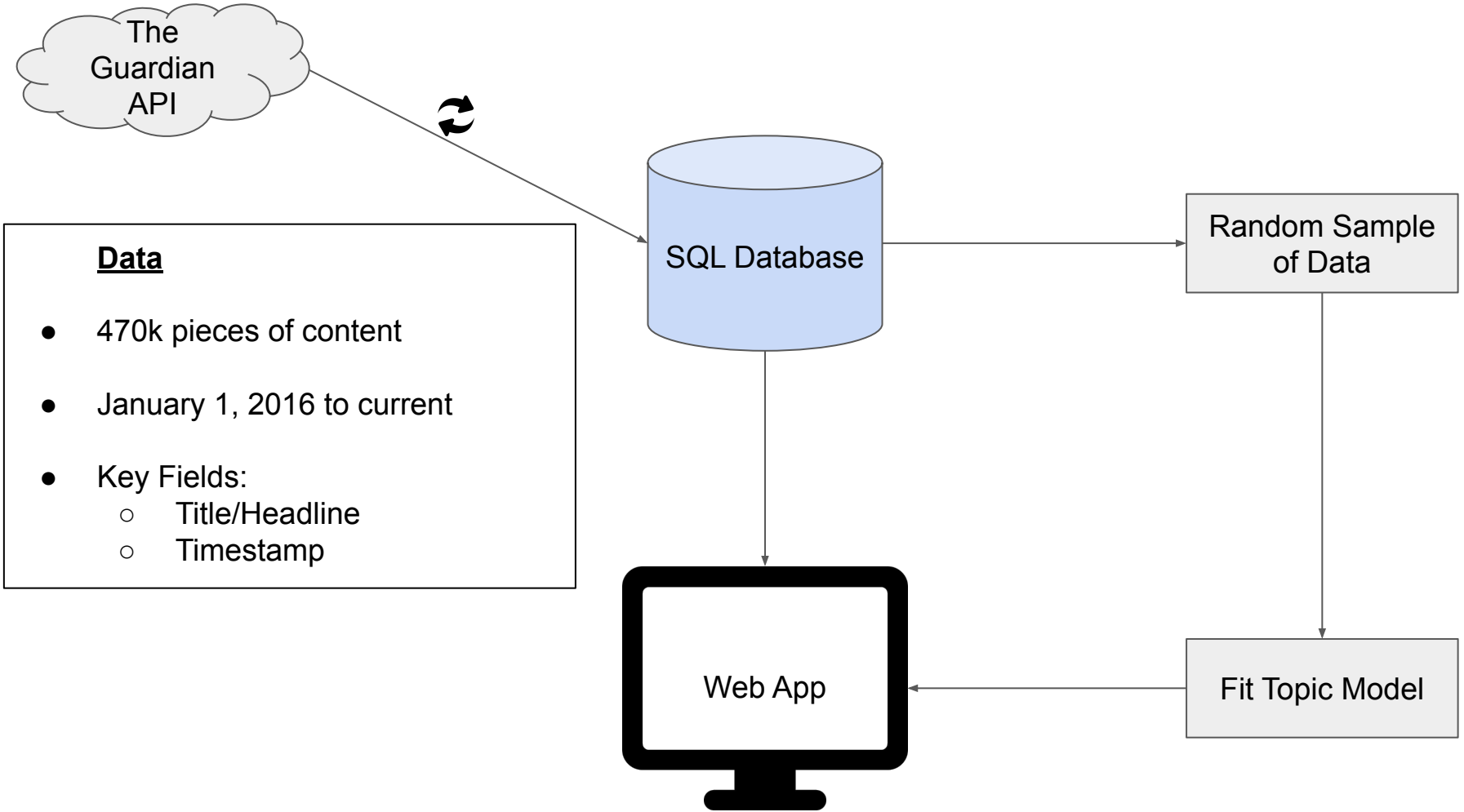
- Dashboard to track publishing volume from The Guardian
- Update database as new content is published
- Make data accessible to a non technical user

News Pulse

1. Data Pipeline
2. NLP & Topic Modeling
3. Web Application

News Pulse

1. **Data Pipeline**
2. NLP & Topic Modeling
3. Web Application



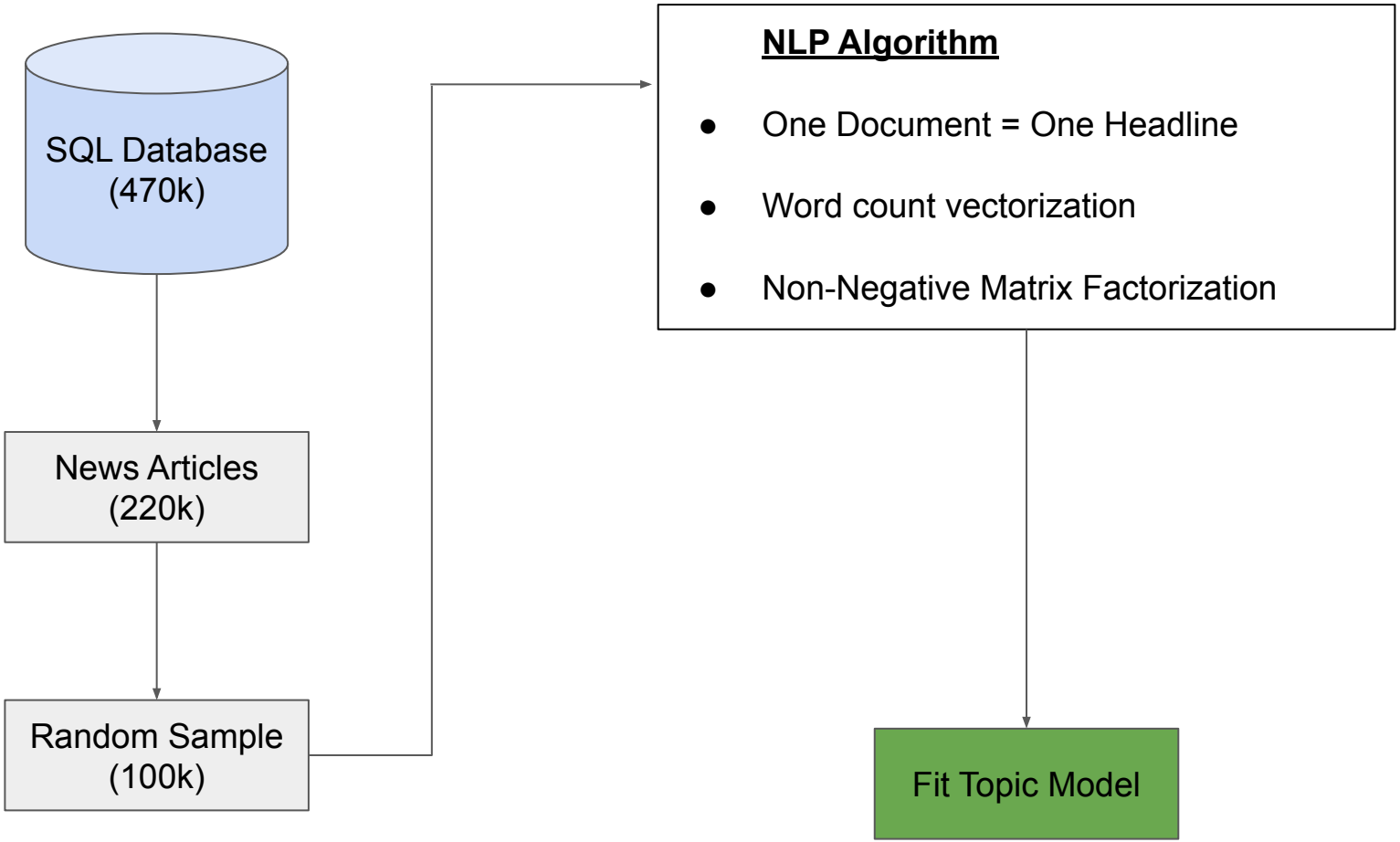
Data

- 470k pieces of content
- January 1, 2016 to current
- Key Fields:
  - Title/Headline
  - Timestamp

News Pulse

1. Data Pipeline
2. **NLP & Topic Modeling**
3. Web Application

NLP: Categorize Content by Title

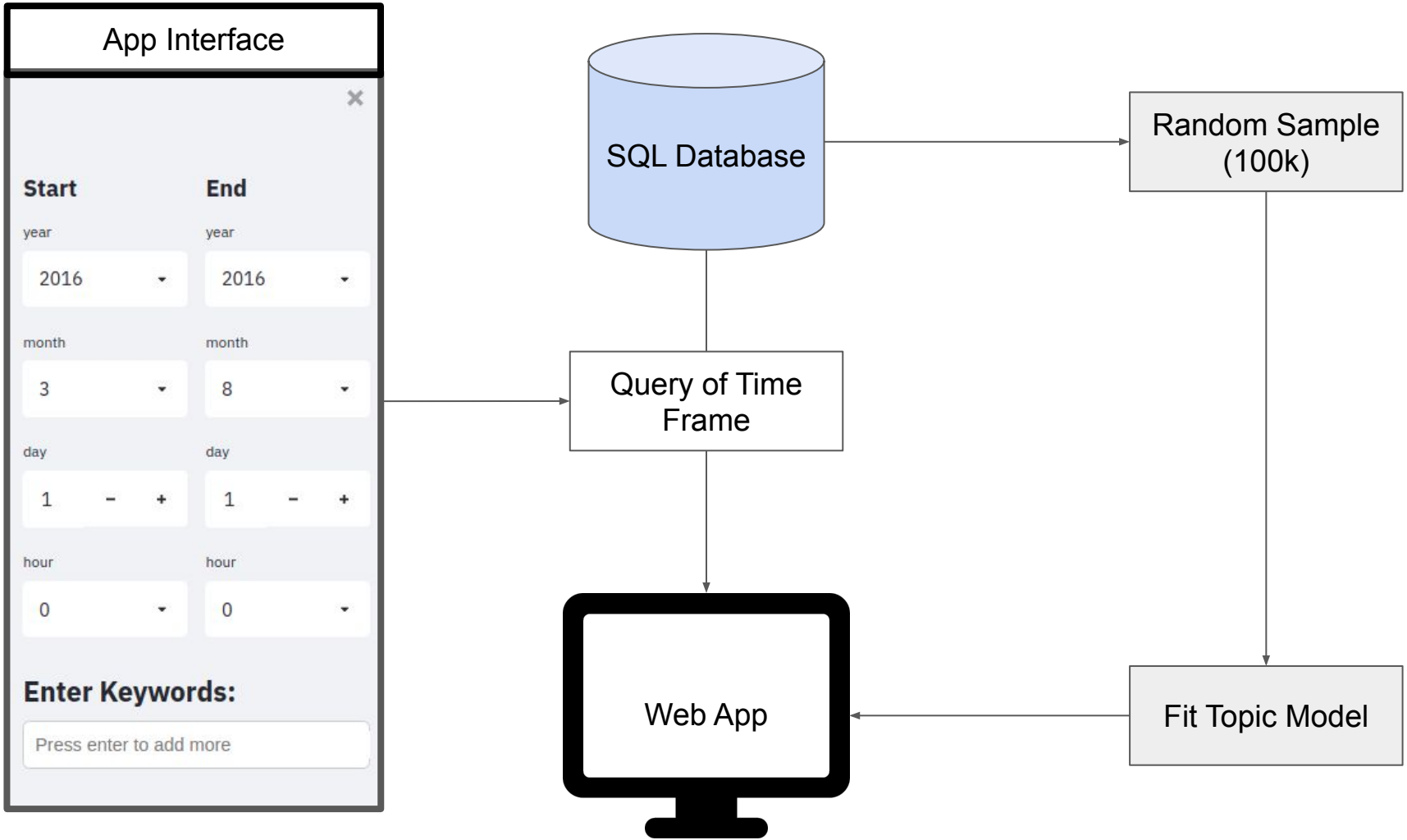

SQL Database
(470k)
News Articles
(220k)
Random Sample
(100k)
NLP Algorithm
One Document = One Headline
Word count vectorization
Non-Negative Matrix Factorization
Fit Topic Model

Topics

1. UK News
2. US News
3. Brexit
4. Letters
5. Crime
6. Climate
7. Children
8. Trump
9. EU Politics
10. Global Politics

News Pulse

1. Data Pipeline
2. NLP & Topic Modeling
3. **Web Application**


App Interface
Start
End
year
2016
year
2016
month
3
month
8
day
1
day
1
hour
0
hour
0
Enter Keywords:
Press enter to add more
SQL Database
Random Sample (100k)
Query of Time Frame
Web App
Fit Topic Model

Start

year: 2016

month: 3

day: 1

hour: 0

End

year: 2016

month: 8

day: 1

hour: 0

Enter Keywords:

Press enter to add more

The Guardian News Cycle Dashboard

Frequency

Weekly

Recent Stories

2016-03-31 23:50:13 Love wins: Guardian marriage equality forum feels the passion...

2016-03-31 23:44:29 Authorities in Monaco raid oil firm HQ in corruption investig...

2016-03-31 23:01:02 Public awareness of link between alcohol and cancer 'worrying...

2016-03-31 23:01:02 UK aid budget swelling coffers of private firms, claims campa...

2016-03-31 23:01:02 Ministers 'letting down mentally ill people' by not collectin...

2016-03-31 23:01:02 Exclusive: royal family considering dramatic Brexit intervent...

2016-03-31 22:21:35 California debates banning long-term solitary confinement for...

2016-03-31 22:11:29 Canada's Prince Edward Island ends abortion ban in province

2016-03-31 21:59:14 Bill Cosby sexual assault allegations to be included in Smith...

UK News

UK | JOB | FIRST | MINISTER | RISK | COURT | TIME | DEATH | REPORT

Clas Ohlson store closures puts dozens of UK jobs at risk

UK courts chief to leave job after just over a year

Spectacular bearded vulture spotted for first time in UK

UK job numbers boosted by women delaying retirement

Guardian editor: EU vote is first time most of UK press is against David Cameron

Show Chart +

US News

US | ELECTION | DEATH | CASE | DEAL | REPORT | JOB | COURT | FIRST

'Bombshell' jobs report puts economic growth back in US election spotlight

Deal or no deal, Yahoo is just the start for the Daily Mail's US push

US election 2016: Republican party nominates Donald Trump for president

Kevin Spacey: House of Cards 'doesn't go far enough' in reflecting US election

US election 'veepstakes': a typical frenzy in lead-up to running mate reveal

Show Chart +

Brexit

BREXIT | DEAL | VOTE | MINISTER | RISK | JOB | CALL | LEADER | REPORT | TIME

Deal or no deal? Brexit and the allure of self-expression

Tata Steel near deal to save Port Talbot plant despite Brexit vote

French border deal won't be affected by Brexit, says Paris

'Stampeding backwards': US officials in Northern Ireland peace deal decry Brexit

Start

year: 2016

month: 3

day: 1

hour: 0

End

year: 2016

month: 8

day: 1

hour: 0

Enter Keywords:

Press enter to add more

2016-03-31 23:01:02 Public awareness of link between alcohol and cancer 'worrying...

2016-03-31 23:01:02 UK aid budget swelling coffers of private firms, claims campa...

2016-03-31 23:01:02 Ministers 'letting down mentally ill people' by not collectin...

2016-03-31 23:01:02 Exclusive: royal family considering dramatic Brexit intervent...

2016-03-31 22:21:35 California debates banning long-term solitary confinement for...

2016-03-31 22:11:29 Canada's Prince Edward Island ends abortion ban in province

2016-03-31 21:59:14 Bill Cosby sexual assault allegations to be included in Smith...

2016-03-31 20:59:09 Think outside the box: San Francisco's horrible, no good, ver...

2016-03-31 20:30:05 My first sighting of this unpromising year

2016-03-31 20:03:59 Donald Trump's most recent attacks on women point to a histor...

2016-03-31 20:00:04 Corrections and clarifications

2016-03-31 19:09:09 BMA launches legal action against junior doctors' contract

2016-03-31 18:58:05 Steel crisis: builders, manufacturers and carmakers urged to ...

2016-03-31 18:55:38 Canadian party leader Thomas Mulcair calls Donald Trump a 'fa...

2016-03-31 18:51:43 Steel craft workers specialise to survive

2016-03-31 18:35:26 Financial traders clash over plans to build masts higher than...

2016-03-31 18:32:42 Pru pay: executives live by different rules

2016-03-31 18:30:42 Brazilian billionaire and owner of the Gherkin charged in bri...

2016-03-31 18:30:29 Port Talbot steelworks closure would be fatal short-termism |...

2016-03-31 18:30:20 A short history of food and antisemitism | Brief letters

2016-03-31 18:29:28 Don't be smug about Trump and abortion | Letters

2016-03-31 18:28:40 Pakistan is winning the war against terrorism | Letter from M...

2016-03-31 18:08:20 Public ownership of Tata Steel could work – just look at Roll...

2016-03-31 18:04:20 Tui Travel reports strong UK demand despite terrorism and wea...

2016-03-31 18:00:38 Zika structure discovered, raising hopes for new ways to comb...

2016-03-31 17:49:00 Polish prime minister favours total ban on abortion

Show Chart +

Brexit

BREXIT | DEAL | VOTE | MINISTER | RISK | JOB | CALL | LEADER | REPORT | TIME

Deal or no deal? Brexit and the allure of self-expression

Tata Steel near deal to save Port Talbot plant despite Brexit vote

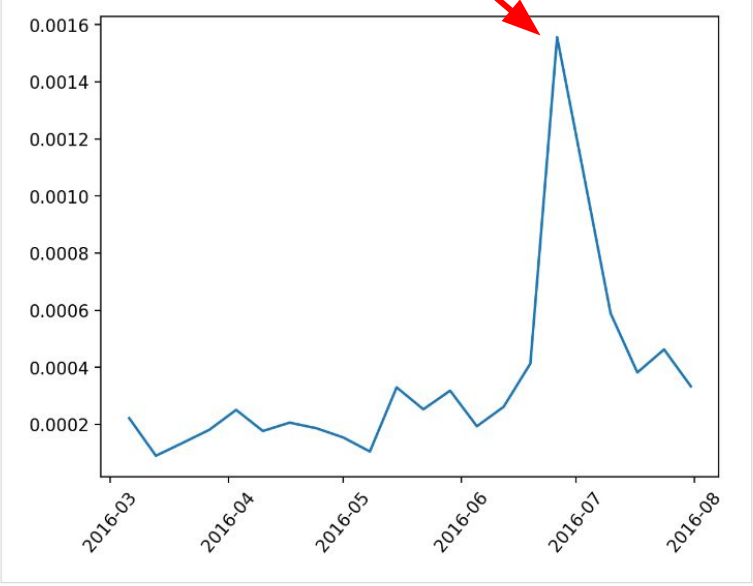
French border deal won't be affecte

'Stampeding backwards': US officia

'Lasses' jobs' replacing industry led to Brexit vote, says clergyman

Brexit Referendum

Show Chart −



Letters

Start

year: 2020

month: 3

day: 2

hour: 0

End

year: 2021

month: 2

day: 1

hour: 0

Enter Keywords:

Press enter to add more

The Guardian News Cycle Dashboard

Frequency

Weekly

Recent Stories

2020-03-31 23:30:34 Whole Foods workers hold 'sick-out' to demand hazard pay duri...

2020-03-31 23:18:47 'Left with nothing': Australia's migrant workforce face desti...

2020-03-31 23:05:05 Coronavirus latest: 31 March at a glance

2020-03-31 23:01:05 Together with porpoise: dolphins team up to end myth of human...

2020-03-31 23:01:05 'We drew up the plan over a brew' - inside operation Nighting...

2020-03-31 22:30:15 UK banks agree to scrap £8bn dividends amid recession fears

2020-03-31 22:12:26 Mexican journalist gunned down in first fatal attack on press...

2020-03-31 21:36:24 Italy sets aside €400m for food vouchers as social unrest mou...

2020-03-31 21:10:27 New York's Andrew Cuomo decries 'eBay'-style bidding war for ...

UK News

UK | CORONAVIRUS | JOB | FIRST | MINISTER | RISK | COURT | TIME | DEATH | REPORT

BLM UK gains legal status and renames as Black Liberation Movement UK

UK watchdog vows to get tough with UK firms over stalled refunds

Wales could lift coronavirus lockdown before rest of UK, says first minister

Fears over coronavirus risk in prisons as first UK inmate case confirmed

UK treasury minister calls for return to work to help economy recover from coronavirus

Show Chart +

US News

US | ELECTION | CORONAVIRUS | DEATH | CASE | DEAL | REPORT | JOB | COURT | FIRST

US election briefing for Australia: Clashing Trump-Biden events highlight US divide

US passes 150,000 coronavirus deaths amid fresh surge in cases

US briefing: California coronavirus death, Super Tuesday fallout and abortion case

US passes 8m coronavirus cases as death toll approaches 220,000

Global report: Donald Trump calls 200,000 US coronavirus deaths 'a shame'

Show Chart +

Brexit

BREXIT | DEAL | VOTE | MINISTER | RISK | JOB | CALL | LEADER | REPORT | TIME

Deal or no deal: how life will look for key industries after Brexit

Brexit: what happens to the UK on 31 December, deal or no deal?

From leave vote to last-ditch deal – a big Brexit timeline

Starmer faces high-profile Labour rebellion before Brexit deal vote

Start

year: 2020

month: 3

day: 2

hour: 0

End

year: 2021

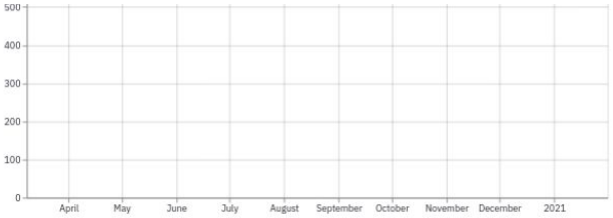
month: 2

day: 1

hour: 0

Enter Keywords:

Press enter to add more



Recent Stories

2020-03-31 23:30:34 Whole Foods workers hold 'sick-out' to demand hazard pay duri...

2020-03-31 23:18:47 'Left with nothing': Australia's migrant workforce face desti...

2020-03-31 23:05:05 Coronavirus latest: 31 March at a glance

2020-03-31 23:01:05 Together with porpoise: dolphins team up to end myth of human...

2020-03-31 23:01:05 'We drew up the plan over a brew' - inside operation Nighting...

2020-03-31 22:30:15 UK banks agree to scrap £8bn dividends amid recession fears

2020-03-31 22:12:26 Mexican journalist gunned down in first fatal attack on press...

2020-03-31 21:36:24 Italy sets aside €400m for food vouchers as social unrest mou...

2020-03-31 21:10:27 New York's Andrew Cuomo decries 'eBay'-style bidding war for ...

2020-03-31 21:00:02 Coronavirus in the Pacific: weekly briefing

2020-03-31 20:30:02 Terrawatch: plastic-rich canyons forming in the deep ocean

2020-03-31 20:24:47 Seven of Donald Trump's most misleading coronavirus claims

2020-03-31 20:13:46 Morning mail: 'stay at home' measures explained, 10% of cases...

2020-03-31 20:00:01 Corrections and clarifications

2020-03-31 20:00:01 Rio Tinto accused of violating human rights in Bougainville f...

2020-03-31 19:40:06 UK care home staff 'at breaking point' as coronavirus cases r...

2020-03-31 19:14:26 NHS staff 'gagged' over coronavirus shortages

2020-03-31 19:12:27 A Govester briefing is proof we're scraping the trustworthine

UK treasury minister calls for return to work to help economy recover from coronavirus

Show Chart +

US News

US | ELECTION | CORONAVIRUS | DEATH | CASE | DEAL | R

US election briefing for Australia: Clashing Trump-Biden ev

US passes 150,000 coronavirus deaths amid fresh surge in cases

US briefing: California coronavirus death, Super Tuesday fallout and abortion case

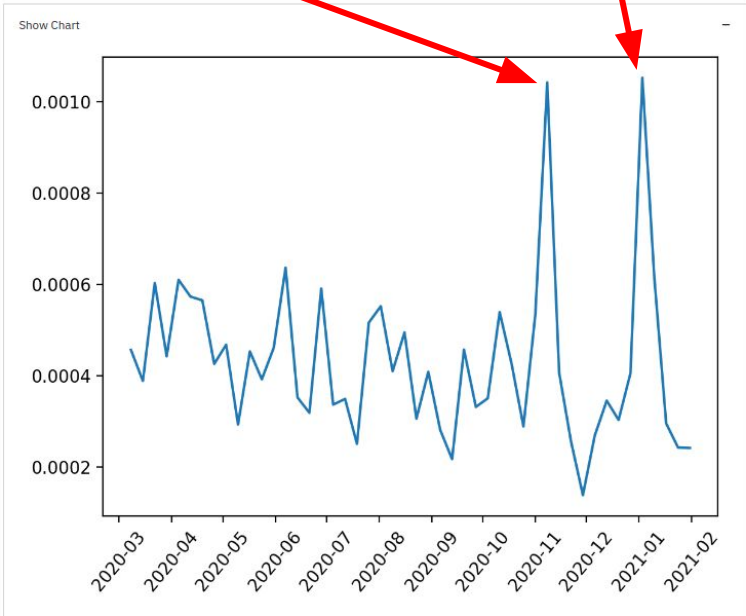
US passes 8m

Global report:

Attack on the Capitol

Biden wins election

Show Chart −



Conclusion

- Automatically get new data as it is created
- Successfully categorized headlines
- Model only needs occasional retraining
- Useful, clean, easy to use web application

Thank you