

Extracting References from Political Speech Auto-Transcripts

Brandon Roberts



THE AUSTIN CUT

"theodore roosevelt"

Score: 10.9005232074

"anthony weiner"

Score: 9.73956778813

"planned parenthood"


Score: 9.36007816643

"saddam hussein"

Score: 9.36007816643

"citibank, goldman sachs"

Score: 9.20544873373



... topic size is a reasonable
predictor of topic quality.
Although there is some overlap,
“bad” topics are generally smaller
than “good” topics.

Mimno et al. (2011)

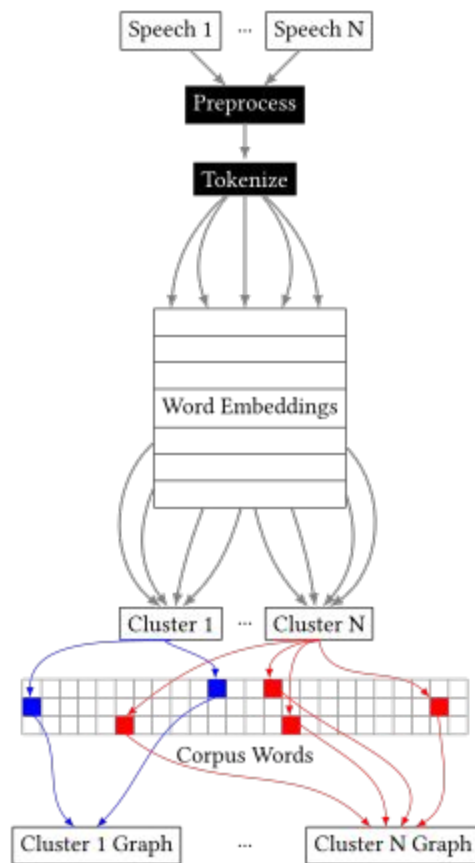
LDA: High Number of Topics (Nonsensical)

| | |
|---|-------|
| morrow, fidelis | 12.13 |
| classifying, doke | 11.15 |
| inputs, textures | 11.03 |
| differ, fended | 10.86 |
| joey, thaw, suk, appreciate, iron, may, even | 3.06 |
| dosomething, simultaneously, stronghold, raise, likely | 2.94 |
| battleship, uss, ponders | 2.92 |
| summit, asian, inches | 2.88 |
| ported, variety, drive | 2.88 |
| rental, fordham, tahoe, isaiah, hudson, celebrity, | |
| polls, buddies, later, iceskating | 0.994 |
| plaza, generators, secondly, muslims, race, particular, not | 0.996 |
| stripes, whose, moment, watched, got | 0.993 |

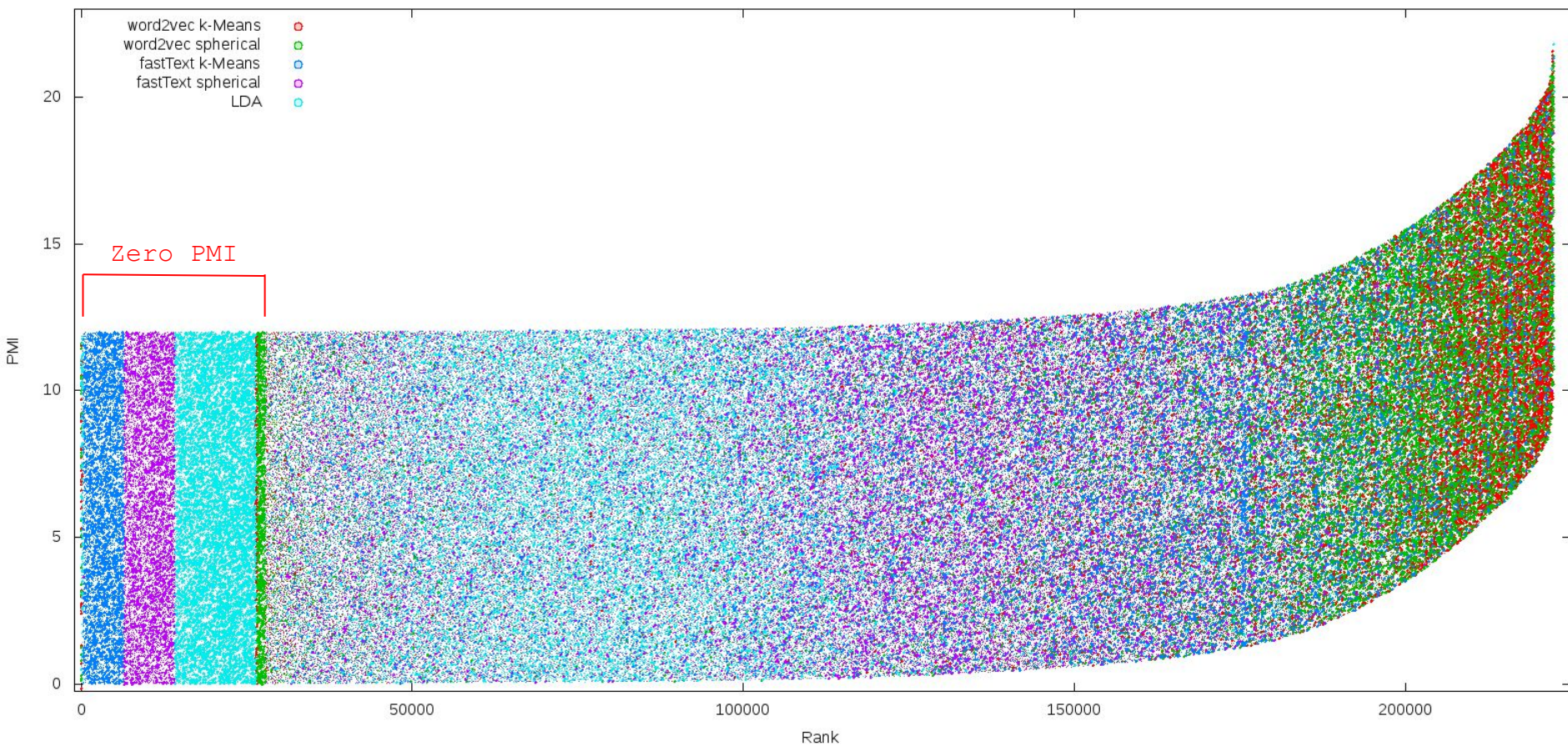
said we will not pay for that f bomb wall the **fm** use the f on live television on cnn i said whoa he's going to be always oh it's going to be so bad tomorrow when you see nobody write about can you imagine if i use the f bomb with that **p** that would be the electric chair right there give trump the electric chair okay if i use the f **pump** so he said we will not under any circumstances always said he was going crazy sitting in a chair that he's gonna have to you know look i wants it but he was going crazy but we will not pay for that f bomb wall and i said well is he in trouble but then i said to myself hey i'm happy because two years ago he says we will never allow that wall to be built on an **hour** at least he knows it's going to be built and the reason look the reason this is very important the reason that taking **a pay** for it we have a great relationship with mexico much better than we have now we don't have a good relationship when sergeant **temporary** see remember sergeant **amerasian** he made a wrong turn he had a lot of difficulties when sergeant **amber** issue is that they kept him in the worst prison they treated him horribly they should let him go he had difficulties they should let him go and they didn't let him go and i don't even know if anybody called i tell you what i would have been on the phone the first night i would call the president of **mexico** i say let him go let him go i don't think obama because he **got was** indifferent you know a long time and i help them out and **grant to help** them out i have to say greta greta van susteren help them out and i help them out by **relationship** as you call him up and say do me a favor guy made a mistake let them out because the people are going crazy and how about this **bergdorf** right sergeant **factor** sergeant here's the way we deal folks we get **bergdahl** and total **trader** we're probably five or six people were killed trying to find him probably definitely five probably six killed going out trying to get him back so we get **bergdahl** and they get five of their worst killers that they've been after and trying to get out of prison for nine years so we get one they get five those five now i saw just the other night they're back in the field trying to kill everybody including us folks folks we got to be so much tougher and so much smarter i call him the president and how about the deal with **around** we get nothing we gave them a hundred and fifty billion dollars and it was only after that remember they took our sailors **are** ten **cellars** and they would have kept him forever except they wanted the money was due the following day right so stupid they're not so they humiliated us **the** humiliated the **seller's** then they let **him** go and i remember after a few days and kerry said we would like to **thank around** to get to **hunt** they got a hundred and fifty billion dollars all right who the hell would have kept them so they want to **captain** believe me and that should have never happened we had hostages we should never have negotiated while those hostages were **help** we should have said **got** into the room they **releasey** our hostages if you don't release the hostages the deals out get up walk out there great negotiators the **radiance** the **persians** a great negotiator is always **happen** you know you're not supposed to generalize i'll generalize ok they've always been great negotiators and we've got a guy who doesn't know what the hell he's doing secretary kerry president obama president obama president obama is incompetent ok he's incompetent remember that he's incompetent now yeah we have a president who is that **doesn't it just doesn't** happen he doesn't have it so here's what's gonna happen here's what's gonna happen i want **you to happen** that you get up and you walk the art of the deal i mean you know give me a break we never walked once you know did you ever see that negotiation we would like to have the following the americans no they

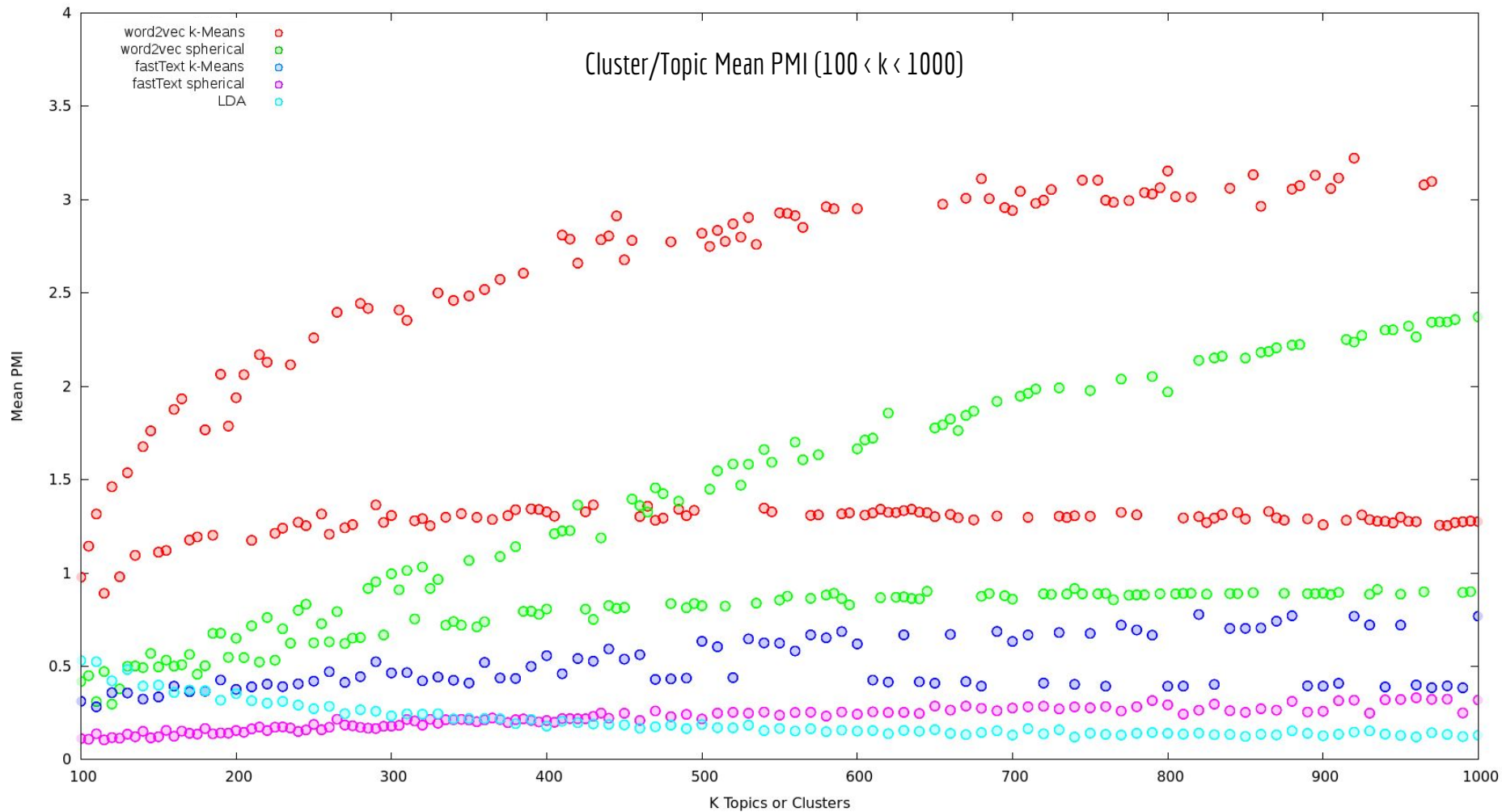
High Level Overview

1. Scrape speech auto-transcripts (YouTube)
2. Preprocess
3. Build Embedding Model
4. Cluster Unique Vectors
5. Identify References
6. Find References in Documents (Dated)

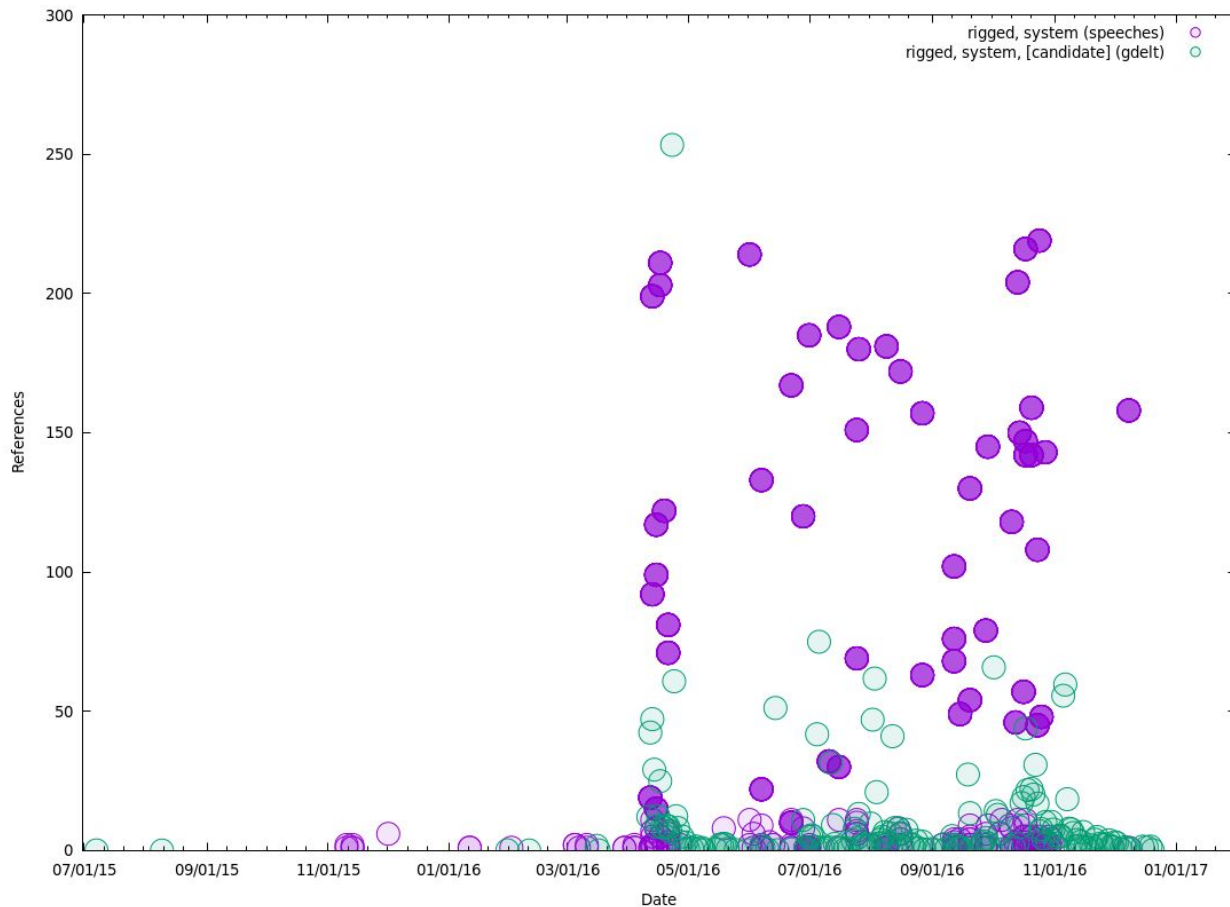


Cluster/Topic PMI Distribution ($100 < k < 1000$)

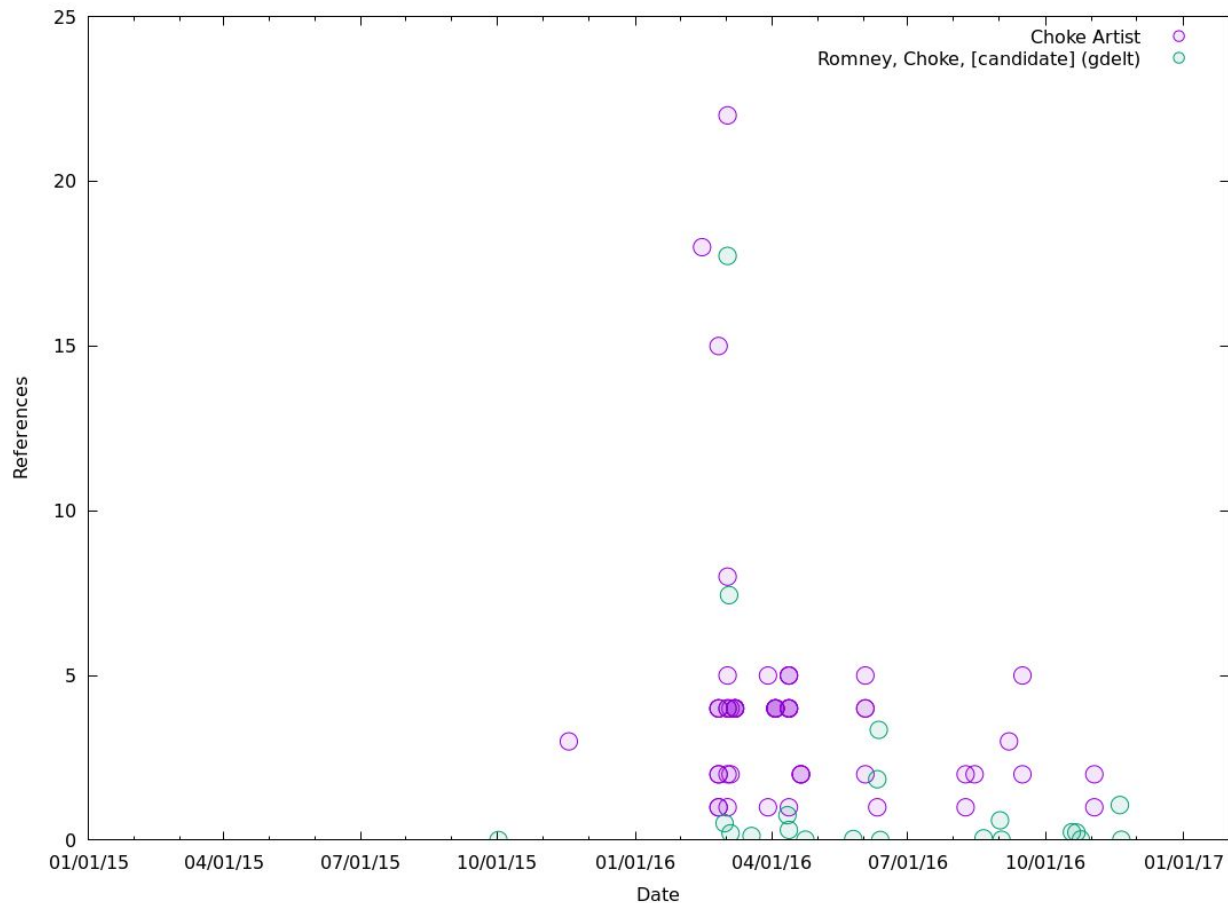




Time Series Extracted References vs. News References (GDelt) - “Rigged System”

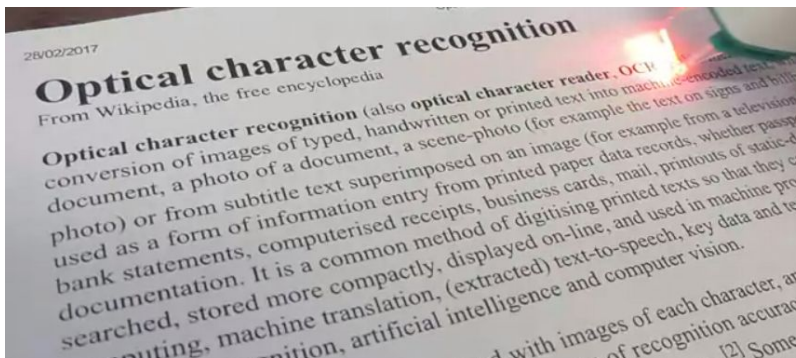


Time Series Extracted References vs. News References (GDelt) - “Romney Choke Artist”



Future Directions

1. More robust evaluation methods
2. Automatic optimization *Lau et al. (2014)*
3. Other inaccurate data forms



Questions?