

Case Study | Creating a corpus of African digital news

Dr. Dani Madrid-Morales
Assistant Professor, University of Houston

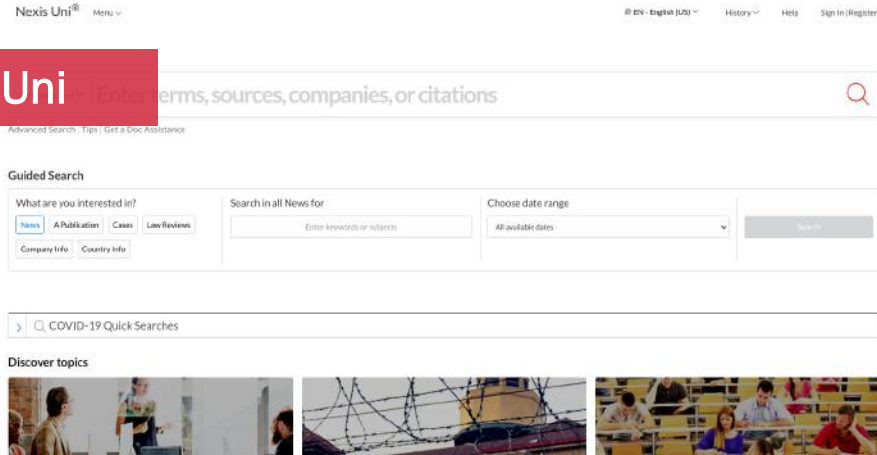
February 25, 2021

Context

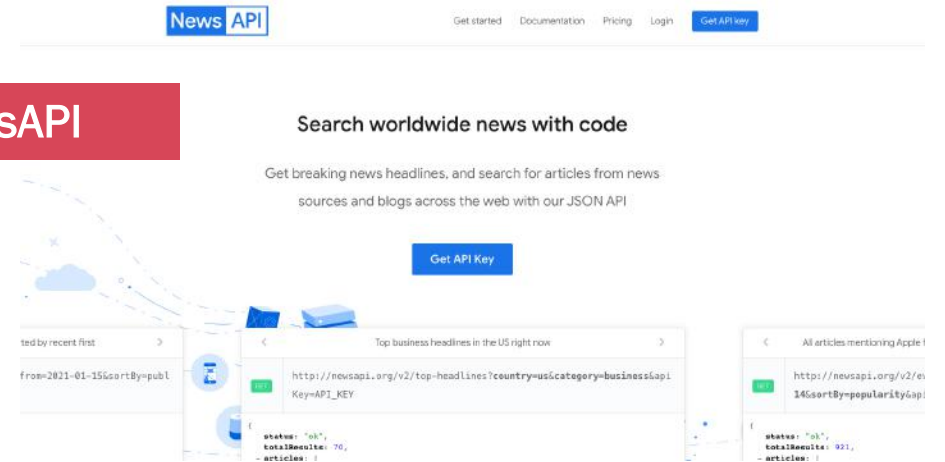
- None of these approaches is superior to other research methodologies. These are useful to answer certain types of questions, but not others.
 - Computational methods are great for **mechanic and repetitive tasks**
 - Computational methods can help **find patterns and classify documents**
 - Computational methods cannot **interpret meanings** of texts
 - Computational methods are **no substitute** of any other research method
- Working with digital news content from African sources can be difficult with existing tools.
 - There are barely no APIs.
 - There are limited efforts to store this kind of data.

Limitations of existing archives

Nexis Uni



NewsAPI



GDELT



Algorithmic biases

Google
Tigray

Q All News Maps Images Videos More Settings Tools

About 1,900,000 results (0.52 seconds)

- Yahoo News**
Ethiopia's Tigray crisis: 'I lost my hand when a soldier tried to rape me'
A teenager caught up in Ethiopia's Tigray conflict tells of her and her grandfather's near-death encounter.
23 hours ago
- Fox News**
Ethiopia's secret war in Tigray region: Ethnic killings, rapes, near-starvation reported
Many women have "conclusively and without a doubt" been raped in the Tigray region, home to Ethiopia's secretive conflict - which may have ...
2 days ago
- BBC News**
Viewpoint: From Ethiopia's Tigray region to Yemen, the dilemma of declaring a famine
Every day, more and more reports of starvation trickle out of the Tigray region of Ethiopia that has been hit by conflict. On Wednesday, Mark ...
1 week ago
- Human Rights Watch**
Ethiopia: Unlawful Shelling of Tigray Urban Areas
Ethiopian federal forces carried out apparently indiscriminate shelling of urban areas in the Tigray region in November 2020 in violation of the ...
5 days ago

Stay up to date on results for **Tigray**. Create alert

Google >
1 2 Next

Google
site:addisstandard.com Tigray

Q All News Maps Images Videos More Settings Tools

When you refine a search with a "site:" operator, Google shows results from all the indexed pages from the website you specified.

About 1,240 results (0.50 seconds)

- addisstandard.com**
News: Ethiopia "deeply concerned" by EU statement on Tigray, stays silent on accusations of Eritrea's role in the...
News: Ethiopia "deeply concerned" by EU statement on Tigray, stays silent on accusations of Eritrea's role in the armed conflict. addisstandard / ...
6 days ago
- addisstandard.com**
News Alert: EU joins the US in calling withdrawal of Eritrean troops from Ethiopia, says they are "fueling the...
EU's latest call on the need to withdraw Eritrean forces from Tigray shows that there is a transatlantic consensus between European Union and ...
1 week ago
- addisstandard.com**
News: Security forces shot dead Tigray TV journalist, his friend in Mekelle; officials decline to comment
Addis Abeba, January 21/2020 – Dawit Kebede Araya, a journalist working for the regional Tigray TV, currently being administered by the ...
4 weeks ago
- addisstandard.com**
News Analysis: Ethiopia Rights Commission latest report on Tigray details killings of civilians, widespread looting...
EHRC further stated that the findings of the Commission's on-going investigation missions in Mekelle other parts of Tigray region since January ...
4 weeks ago

Stay up to date on results for **site:addisstandard.com Tigray**. Create alert

Google >
1 2 3 Next

Africa Digital News Corpus

- In 2019 we started a pilot project to monitor news content on digital news in Africa in real time to create a **live corpus of news**.

- We built this using R and Python.

<https://doi.org/10.18738/T8/UKJZ3E>

African Digital News Dataverse (University of Houston)

Texas Data Repository Dataverse > University of Houston Dataverse > African Digital News Dataverse >

Corpus of African Digital News from 1,200 Websites Formatted for Text Mining / Computational Text Analysis

Version 1.0

Madrid-Morales, Dani; Lindner, Peggy; Periyasamy, Madhumitha, 2021, "Corpus of African Digital News from 1,200 Websites Formatted for Text Mining / Computational Text Analysis", <https://doi.org/10.18738/T8/UKJZ3E>, Texas Data Repository Dataverse, V1

Cite Dataset - Learn about Data Citation Standards.

Access Dataset - Contact Owner Share

Dataset Metrics 1 View 0 Downloads 0 Citations

Description

This dataset includes a corpus 3 million news articles published by 1,200 African news organizations between December 6, 2020 and January 4, 2021. The texts have been pre-processed (punctuation and English stopwords have been removed, features have been lowercased, lemmatized and POS-tagged) and stored in commonly used formats for text mining/computational text analysis. Users are advised to read the documentation for an explanation of the data collection process.

This dataset includes the following items:

- A long table of lowercased and lemmatized tokens with the following additional variables: POS tags, document id, sentence id, token id (stored as a tibble).
- A document-feature (DFM) matrix with raw counts of feature frequencies in each news item (stored as a quanteda dfm object). The DFM comes with the following metadata for each news item: date of publication and source URL.
- A document-feature (DFM) matrix that uses the tf-idf statistic to weight the frequency of features in each news item (stored as a quanteda dfm object). The DFM comes with the following metadata for each news item: date of publication and source URL.

All items are stored in formats readable in R. Consult the documentation provided with the data on how to load

Subject Arts and Humanities; Social Sciences

Keyword news, Africa, digital news, text mining, computational

Open Access
Open Source

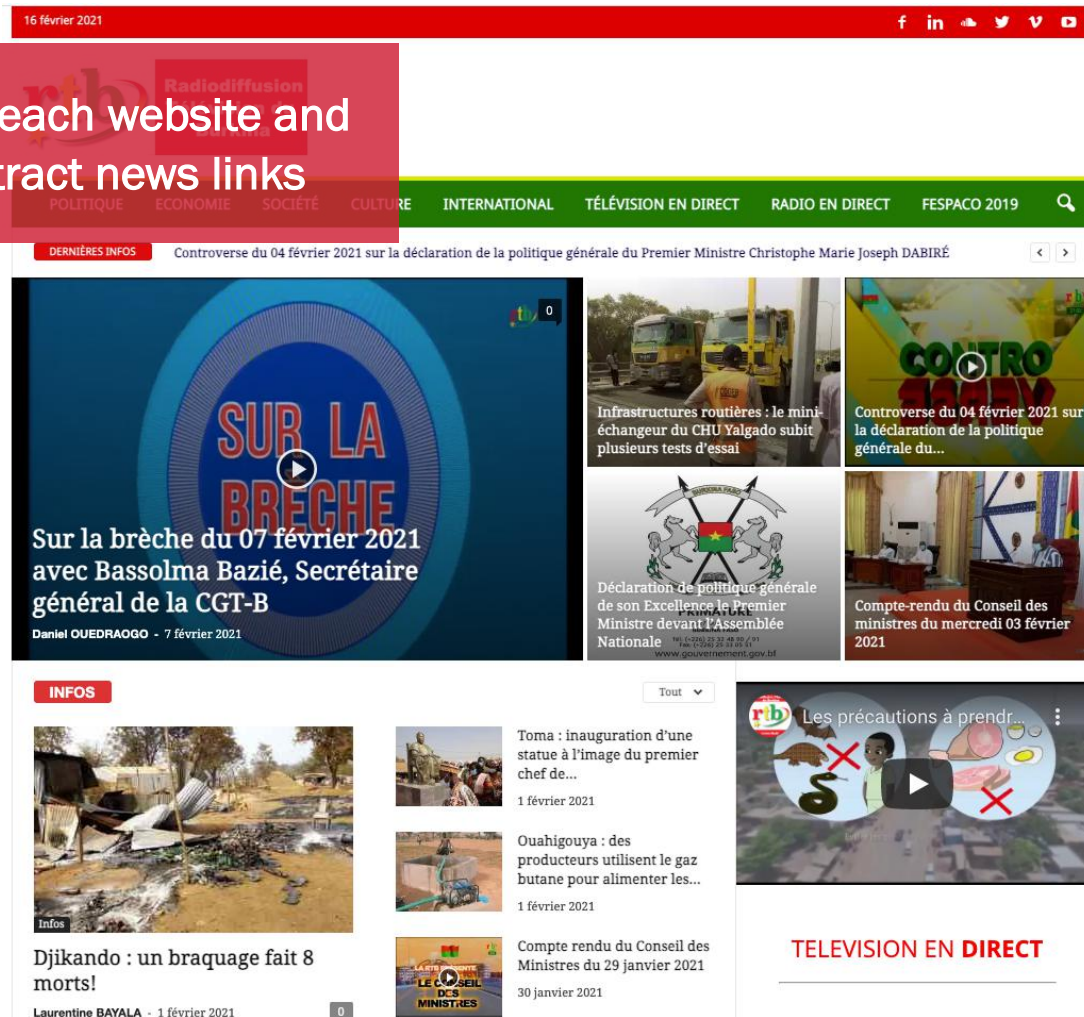
Process – Step 1: Seed Links

Create a list of pages you want to scrape

A	B	C	D	E	F	G
country_name	country_is	medium_name	alexa_scd	source	url	news_site
Burkina Faso	BF	http://www.aib.media	NA	xyzlinks	http://www.aib.media	TRUE
Burkina Faso	BF	http://www.aouaga.com	NA	xyzlinks	http://www.aouaga.com	TRUE
Burkina Faso	BF	http://www.aujourd8.net	NA	xyzlinks	http://www.aujourd8.net	TRUE
Burkina Faso	BF	http://www.burkina24.com	NA	xyzlinks	http://www.burkina24.com	TRUE
Burkina Faso	BF	http://www.burkinainfo.com	NA	xyzlinks	http://www.burkinainfo.com	TRUE
Burkina Faso	BF	http://www.faso-actu.net	NA	xyzlinks	http://www.faso-actu.net	TRUE
Burkina Faso	BF	http://www.fasopresse.net	NA	xyzlinks	http://www.fasopresse.net	TRUE
Burkina Faso	BF	http://www.fasozine.com	NA	xyzlinks	http://www.fasozine.com	TRUE
Burkina Faso	BF	http://www.leconomistedufaso.bf	NA	xyzlinks	http://www.leconomistedufaso.bf	TRUE
Burkina Faso	BF	http://www.lefaso.net	NA	xyzlinks	http://www.lefaso.net	TRUE
Burkina Faso	BF	http://www.lepays.bf	NA	xyzlinks	http://www.lepays.bf	TRUE
Burkina Faso	BF	http://www.lepays.bf	NA	xyzlinks	http://www.lepays.bf	TRUE
Burkina Faso	BF	http://www.lesaffairesbf.com	NA	xyzlinks	http://www.lesaffairesbf.com	TRUE
Burkina Faso	BF	http://www.lexpressdufaso-bf.com	NA	xyzlinks	http://www.lexpressdufaso-bf.com	TRUE
Burkina Faso	BF	http://www.lobserveur.bf	NA	xyzlinks	http://www.lobserveur.bf	TRUE
Burkina Faso	BF	http://www.lobserveur.bf	NA	xyzlinks	http://www.lobserveur.bf	TRUE
Burkina Faso	BF	http://www.monpulsar.com	NA	xyzlinks	http://www.monpulsar.com	TRUE
Burkina Faso	BF	http://www.mutationsbf.info	NA	xyzlinks	http://www.mutationsbf.info	TRUE
Burkina Faso	BF	http://www.omegabf.info	NA	xyzlinks	http://www.omegabf.info	TRUE
Burkina Faso	BF	http://www.ouagafm.bf	NA	xyzlinks	http://www.ouagafm.bf	TRUE
Burkina Faso	BF	http://www.radiob24.com	NA	xyzlinks	http://www.radiob24.com	TRUE
Burkina Faso	BF	http://www.rtb.bf	NA	xyzlinks	http://www.rtb.bf	TRUE
Burkina Faso	BF	http://www.savanefm.bf	NA	xyzlinks	http://www.savanefm.bf	TRUE
Burkina Faso	BF	http://www.sidwaya.info	NA	xyzlinks	http://www.sidwaya.info	TRUE
Burkina Faso	BF	http://www.sidwaya.info	NA	xyzlinks	http://www.sidwaya.info	TRUE
Burundi	BI	http://www.abpinfos.com	NA	xyzlinks	http://www.abpinfos.com	TRUE
Burundi	BI	http://www.arib.info	NA	xyzlinks	http://www.arib.info	TRUE
Burundi	BI	http://www.burundi24.wordpress.com	NA	xyzlinks	http://www.burundi24.wordpress.com	TRUE
Burundi	BI	http://www.burundi-agnews.org	NA	xyzlinks	http://www.burundi-agnews.org	TRUE
Burundi	BI	http://www.burundi-eco.com	NA	xyzlinks	http://www.burundi-eco.com	TRUE
Burundi	BI	http://www.burundi-forum.org	NA	xyzlinks	http://www.burundi-forum.org	TRUE
Burundi	BI	http://www.burundi-information.net	NA	xyzlinks	http://www.burundi-information.net	TRUE
Burundi	BI	http://www.fr.itaraburundi.com	NA	xyzlinks	http://www.fr.itaraburundi.com	TRUE
Burundi	BI	http://www.isanganiro.org	NA	xyzlinks	http://www.isanganiro.org	TRUE
Burundi	BI	http://www.itaraburundi.com	NA	xyzlinks	http://www.itaraburundi.com	TRUE
Burundi	BI	http://www.iwacu-burundi.org	NA	xyzlinks	http://www.iwacu-burundi.org	TRUE

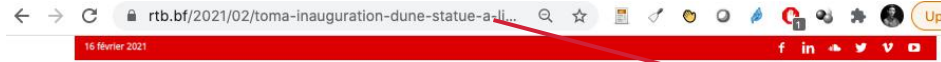
Process – Step 2: Retrieve links

Visit each website and extract news links



Link	URL
Toma : inauguration d'une statue à l'image du premier chef de...	https://www.rtb.bf/2021/02/toma-inauguration-dune-statue-a-limage-du-premier-chef-de-canton/
Ouahigouya : des producteurs utilisent le gaz butane pour alimenter les...	https://www.rtb.bf/2021/02/ouahigouya-des-producteurs-utilisent-le-gaz-butane-pour-alimenter-les-motopompes/
Compte rendu du Conseil des Ministres du 29 janvier 2021	https://www.rtb.bf/2021/01/compte-rendu-du-conseil-des-ministres-du-29-janvier-2021/
Compte-rendu du conseil des ministres du vendredi 29 janvier 2021	https://www.rtb.bf/2021/01/compte-rendu-du-conseil-des-ministres-du-vendredi-29-janvier-2021/

Process – Step 3: Extracting data



Extract data that is relevant for us

Toma : inauguration d'une statue à l'image du premier chef de canton.



Image MCAT

TOUJOURS AU COEUR DES GRANDS ÈVÈNEMENTS

Les autorités municipales de Toma ont procédé ce samedi 30 janvier 2021 au baptême d'un rond-point où trône une statue d'Issa Gue Paré. Cette statue immortalise le tout premier chef de canton de Toma de 1916 à 1943. Pendant 27 ans de règne, Issa Gue Paré a consacré sa vie au rayonnement de son canton dans plusieurs domaines

publish_date	source	country_iso	full_text	web_url
2020-04-07	abidjan.net	CI	Société Hamed Bakayoko déclaré positif au coronavirus...	https://news.abidjan.net/h/671218.html
2020-04-27	abidjan.net	CI	Société COVID 19: Un don en vivres et non vivres offe...	https://news.abidjan.net/h/672003.html
2020-04-20	abidjan.net	CI	Économie Covid-19: le directeur général de IOMS peu...	https://news.abidjan.net/h/671694.html
2020-04-22	abidjan.net	CI	Économie Hausse du coût de la viande : les acteurs in...	https://news.abidjan.net/h/671803.html
2020-04-29	abidjan.net	CI	Société Secteur du tourisme et des loisirs : les résulta...	https://news.abidjan.net/h/672090.html
2020-03-21	abidjan.net	CI	Économie Covid-19: Les mesures urgentes prises par ...	https://news.abidjan.net/h/670625.html
2020-03-23	abidjan.net	CI	Société Journée mondiale de l'eau 2020 toujours d'ac...	https://news.abidjan.net/h/670673.html
2020-03-27	abidjan.net	CI	International New York, nouvel épice de la pandé...	https://news.abidjan.net/h/670803.html
2020-05-06	abidjan.net	CI	Société Ramadan : Hamed Bakayoko offre des vivres a...	https://news.abidjan.net/h/672361.html
2020-05-12	abidjan.net	CI	Société Coronavirus : Le Premier Ministre met la minis...	https://news.abidjan.net/h/672617.html
2020-05-14	abidjan.net	CI	Société Reprise des travaux de la Nationale A3 à Niak...	https://news.abidjan.net/h/672735.html
2020-05-27	abidjan.net	CI	Politique Présidentielle 2020: la révision de la liste éle...	https://news.abidjan.net/h/673528.html
2020-02-27	abidjan.net	CI		https://news.abidjan.net/h/669714.html
2020-04-03	abidjan.net	CI	Économie Covid- 19 Des producteurs de café-cacao r...	https://news.abidjan.net/h/671048.html
2020-04-05	abidjan.net	CI	Société Le préfet de Korhogo appelle la population à s...	https://news.abidjan.net/h/671111.html
2020-04-10	abidjan.net	CI	Région COVID-19 : Agboville lance la phase répressiv...	https://news.abidjan.net/h/671331.html
2020-04-12	abidjan.net	CI	Société Pâques 2020 : L'Eglise méthodiste plaide pour ...	https://news.abidjan.net/h/671409.html
2020-04-15	abidjan.net	CI	Art et Culture Covid 19: Des artistes et journalistes se...	https://news.abidjan.net/h/671480.html
2020-04-10	abidjan.net	CI	Société Coronavirus : Le centre de dépistage de Yopo...	https://news.abidjan.net/h/671348.html
2020-04-04	abidjan.net	CI	Société Covid-19 : Sensibilisation à outrance à l'intéri...	https://news.abidjan.net/h/671076.html
2020-04-09	abidjan.net	CI	Société Covid-19: Voici les techniques utilisées par d...	https://news.abidjan.net/h/671283.html
2020-04-03	abidjan.net	CI	Société Treichville/ Lutte contre le Coronavirus : le co...	https://news.abidjan.net/h/671067.html
2020-04-15	abidjan.net	CI	Société Covid-19: L'État de Côte d'Ivoire proroge le dél...	https://news.abidjan.net/h/671500.html
2020-04-19	abidjan.net	CI	Société Bruno Koné fait des dons aux ressortissants d...	https://news.abidjan.net/h/671672.html
2020-03-06	abidjan.net	CI	Société Social: Report du gala de la Fondation Childre...	https://news.abidjan.net/h/670091.html
2020-03-18	abidjan.net	CI	Société Coronavirus: Latelier de formation des format...	https://news.abidjan.net/h/670503.html
2020-03-19	abidjan.net	CI	Politique Covid-19: La plateforme pro-Gbagbo, EDS, l...	https://news.abidjan.net/h/670561.html
2020-03-17	abidjan.net	CI	Société Crise du Coronavirus : Les Abidjanais vident l...	https://news.abidjan.net/h/670484.html
2020-03-27	abidjan.net	CI	Économie Impacts du Covid-19 dans l'UEMOA : Avis r...	https://news.abidjan.net/h/670824.html
2020-03-26	abidjan.net	CI	Économie Covid-19/Abidjan : Le marché gouro ferme...	https://news.abidjan.net/h/670770.html
2020-05-06	abidjan.net	CI	Société Lutte contre le Covid-19: démarrage officiel d...	https://news.abidjan.net/h/672345.html
2020-05-04	abidjan.net	CI	Société Ramadan : 400 familles démunies soulagées à...	https://news.abidjan.net/h/672295.html
2020-05-10	abidjan.net	CI	Société Lutte contre le Covid-19: les commerçants du...	https://news.abidjan.net/h/672540.html
2020-05-13	abidjan.net	CI	Société Des enseignants musulmans offrent des kits a...	https://news.abidjan.net/h/672674.html
2020-05-12	abidjan.net	CI	Société Covid-19: des équipements sanitaires offerts ...	https://news.abidjan.net/h/672628.html
2020-05-20	abidjan.net	CI	Société Nuit du Qadr: Les musulmans ivoiriens invités...	https://news.abidjan.net/h/673090.html

What does the code look like

```
##Function: Grab links from initial URLs##
get_links <- function(url){

  withTimeout({Sys.sleep(2)

    page <- read_html(url)

    web_link <- page %>%
      html_nodes("a") %>%
      html_attr("href")

    text_links <- page %>%
      html_nodes("a") %>%
      html_text()

    tibble(web_link,
           text_links,
           url)
  }, timeout = 10, onTimeout = "silent")
}
###
```

```
##Function: scrapes out text, date, and raw HTML data from a given link##
get_text_from_py <- function(url_link){

  withTimeout({
    traf <- import("trafilatura")
    requests <- import("requests")
    rand <- import("requests_random_user_agent")
    goose <- import("goose3")

    r <- traf$fetch_url(url_link)
    web_text <- traf$extract(r)
    web_raw1 <- r[1]

    web_date <- traf$metadata$extract_metadata(r) %>%
      str_extract("[0-9]{4}-[0-9]{2}-[0-9]{2}")

    g = goose$Goose({browser_user_agent = browser_agent})

    article <- g$extract(url = url_link)
    web_text2 <- article$cleaned_text
    web_date2 <- as.character(article$publish_datetime_utc)
    web_raw2 <- article$raw_html

    if (is.null(web_date2) == TRUE){
      web_date2 <- ""
    }

    if (is.null(web_text2) == TRUE){
      web_text2 <- ""
    }

    g$close()

    tibble(web_url = url_link,
          web_date = web_date,
          web_date2 = web_date2,
          web_sysdate = Sys.Date(),
          web_text = web_text,
          web_text2 = web_text2,
          web_raw1 = web_raw1,
          web_raw2 = web_raw2)
  }, timeout = 10, onTimeout = "silent")
}
```

Step 4: Processing the data (optional)

- Use spacy (or udpipe) to tokenize (or lemmatize) the text and POS (part of speech tag)
- We can then remove tokens that we don't "need" (stopwords, numbers, punctuation...)

	doc_id	sentence_id	token_id	token	lemma	pos	entity
1	STMID_006086	1	1	When	when	ADV	
2	STMID_006086	1	2	Kem	Kem	PROPN	
3	STMID_006086	1	3	Senou	Senou	PROPN	
4	STMID_006086	1	4	Pavel	Pavel	PROPN	
5	STMID_006086	1	5	Daryl	Daryl	PROPN	
6	STMID_006086	1	6	,	,	PUNCT	
7	STMID_006086	1	7	a	a	DET	
8	STMID_006086	1	8	21-year	21-year	NUM	DATE_B
9	STMID_006086	1	9	-	-	PUNCT	DATE_I
10	STMID_006086	1	10	old	old	ADJ	DATE_I
11	STMID_006086	1	11	Cameroonian	cameroonian	ADJ	NORP_B
12	STMID_006086	1	12	student	student	NOUN	
13	STMID_006086	1	13	living	live	VERB	
14	STMID_006086	1	14	in	in	ADP	
15	STMID_006086	1	15	the	the	DET	
16	STMID_006086	1	16	Chinese	chinese	ADJ	NORP_B
17	STMID_006086	1	17	city	city	NOUN	
18	STMID_006086	1	18	of	of	ADP	
19	STMID_006086	1	19	Jingzhou	Jingzhou	PROPN	GPE_B
20	STMID_006086	1	20	,	,	PUNCT	

Applications

- This same process can be used to collect content from government websites, discussion fora, citizen journalism platforms...
- With this type of data...
 - We can answer **descriptive questions** about content
 - We can answer questions about **linguistic patterns**
 - We can **compare sources** across geographic spaces
 - We can trace **longitudinal patterns**
 - We can **triangulate text data with fieldwork**

Limitations

- This approach would (in principle) exclude data that is behind **paywalls** (think of Daily Maverick in SA or Nation Media in KE); it is internet data heavy.
- These approaches are heavy on the **amount of bandwidth** that is required (cloud computing as alternative).
- Downtime of websites and other technical problems might limit how much data we can collect, so it is important to acknowledge **biases in our dataset**.

Case Study | African digital media reporting on COVID-19 and China

Dr. Dani Madrid-Morales
University of Houston

February 25, 2021

Previous work

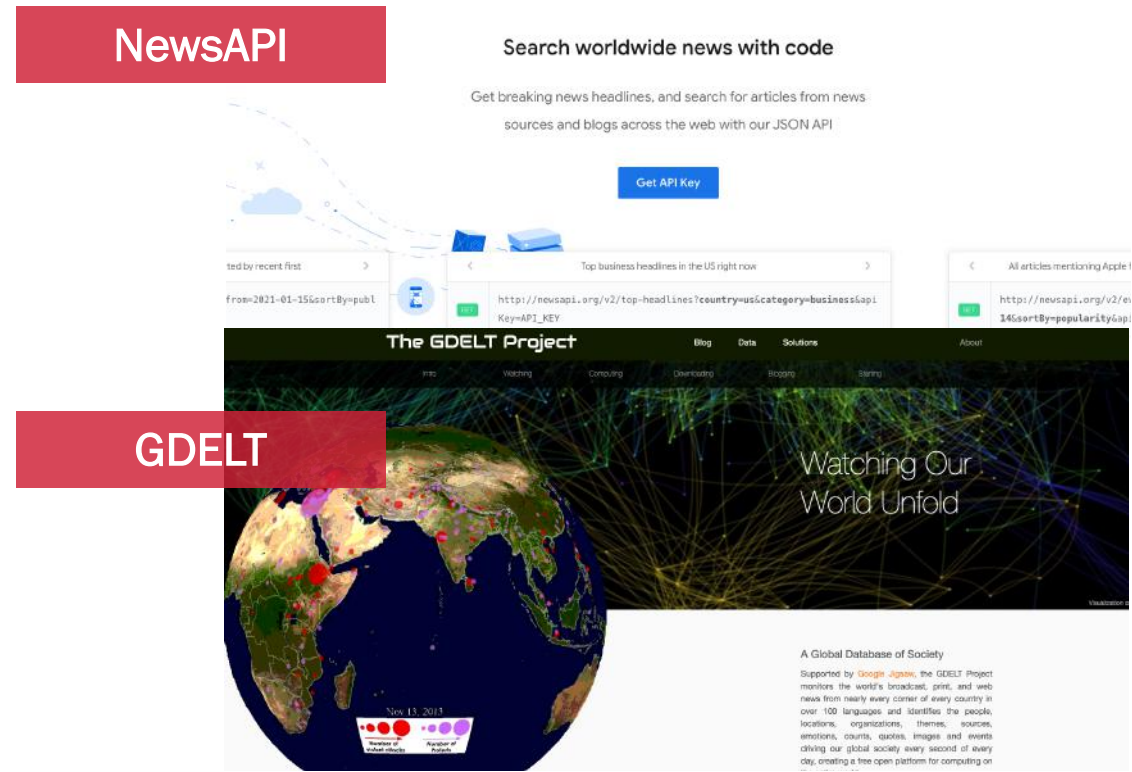
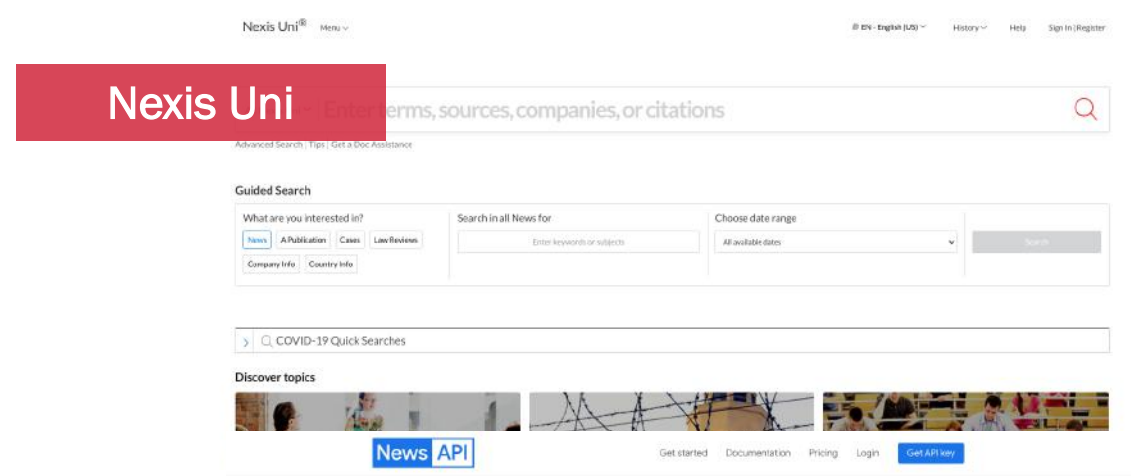
- Interviews with journalists in Kenya and South Africa, and observation of newsroom practices.
 - <https://doi.org/10.1080/23743670.2019.1627230>
- Focus groups with media users to understand how much Chinese news content makes its way to the news diet of audiences in Kenya & South Africa.
 - <http://ijoc.org/index.php/ijoc/article/view/7809/2355>
- Analysis of social media use by Chinese diplomats to shape narratives about China across African countries.
 - <https://hdl.handle.net/10657/6258>

Research Questions

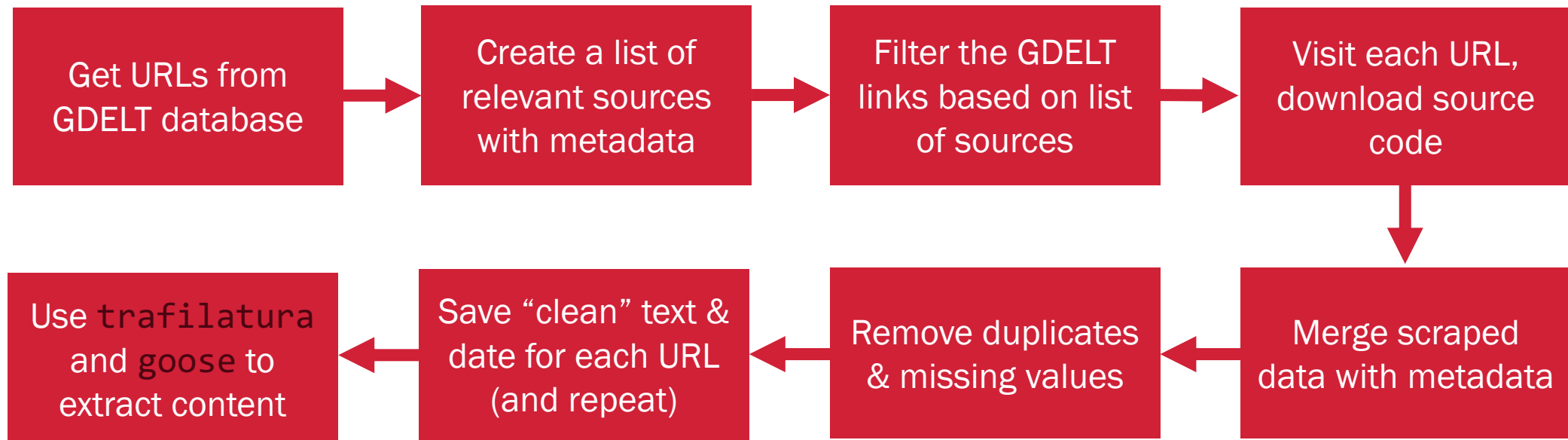
- RQ1: How often did African digital media discuss COVID-19 and China between January and May 2020?
- RQ2: How much of the reporting on African digital media came from foreign sources?
- RQ3: Were there differences in text reuse between French-speaking and English-speaking countries?
- RQ4: How much of the narrative around COVID-19 by China made it to African digital media?

Approach

- Compiled 500,000 news stories from French & English language **digital sources based in Africa & UK, France, US, China.**
- Used a combination of **computational methods** (text similarity algorithms & structural topic modeling)



Data Gathering Workflow



Data Gathering

Downloading
GDELT URLs

← → × Not Secure | data.gdeltproject.org/gdeltv2/masterfilelist.txt

```
150383 297a16b493de7cf6ca809a7cc31d0b93 http://data.gdeltproject.org/gdeltv2/20150218230000.export.CSV.zip
218084 bb27f70ba45f69a17ea6ed7755e9f8ff http://data.gdeltproject.org/gdeltv2/20150218230000.mentions.CSV.zip
10768507 ea8dde0beb0ba98810a92db068c0ce99 http://data.gdeltproject.org/gdeltv2/20150218230000.gkg.csv.zip
...
243548 9568632b94de04c3cbbd627382ccd0abc http://data.gdeltproject.org/gdeltv2/20150219020000.mentions.CSV.zip
```

Source: <https://blog.gdeltproject.org/gdelt-2-0-our-global-world-in-realtime/>

Filtering links by
relevant URLs

```
# Read raw GDELT data
covid_gdelt_files <- list.files(pattern = "COVID", path = getwd()) %>%
  map(function(x){
    temp <- fread(x) %>%
      tibble()
  }) %>%
  bind_rows()

# Save African links in GDELT EN
covid_gdelt_files_AFR <- covid_gdelt_files %>%
  select(web_url = DocumentIdentifier,
         date_sources = GKGRECORDID,
         medium_name = SourceCommonName) %>%
  right_join(african_sources)
fwrite(covid_gdelt_files_AFR, "COVID_GDELT_AFROnly.csv")

# Save African links in GDELT FR
covid_gdelt_files_AFR <- covid_gdelt_files %>%
  select(web_url = DocumentIdentifier,
         date_sources = GKGRECORDID,
         medium_name = SourceCommonName) %>%
  right_join(african_sourcesFR)
fwrite(covid_gdelt_files_AFR, "COVID_GDELT_AFROnly_FR.csv")

# Save Chinese links in GDELT
covid_gdelt_files_CN <- covid_gdelt_files %>%
  select(web_url = DocumentIdentifier,
         date_sources = GKGRECORDID,
         medium_name = SourceCommonName) %>%
  right_join(tibble(medium_name = c("chinadaily.com.cn",
                                   "globaltimes.cn",
                                   "xinhuanet.com",
                                   "chinaview.cn",
                                   "news.cn",
                                   "cgtn.com",
                                   "people.cn")))
fwrite(covid_gdelt_files_CN, "COVID_GDELT_CNOnly.csv")
```

Scraping Text Data

```
get_text_from_py <- function(url_link){  
  withTimeout({  
    traf <- import("trafilatura")  
    requests <- import("requests")  
    rand <- import("requests_random_user_agent")  
    goose <- import("goose3")  
  
    r <- traf$fetch_url(url_link)  
    web_text <- traf$extract(r)  
  
    web_date <- traf$metadata$extract_metadata(r) %>%  
      str_extract("[0-9]{4}-[0-9]{2}-[0-9]{2}")  
  
    g = goose$Goose({browser_user_agent = 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_2)'})  
  
    article <- g$extract(url = url_link)  
    web_text2 <- article$cleaned_text  
    web_date2 <- as.character(article$publish_datetime_utc)  
  
    if (is.null(web_date2) == TRUE){  
      web_date2 <- ""  
    }  
  
    if (is.null(web_text2) == TRUE){  
      web_text2 <- ""  
    }  
  
    g$close()  
  })  
}
```

Import Python libraries

Extract text and dates 1

Extract text and dates 2

Analysis

- Approach 1: Use **word frequencies** to compare texts and find how similar they are
- Approach 2: Use word frequencies (and metadata) to **classify documents** into N topics.

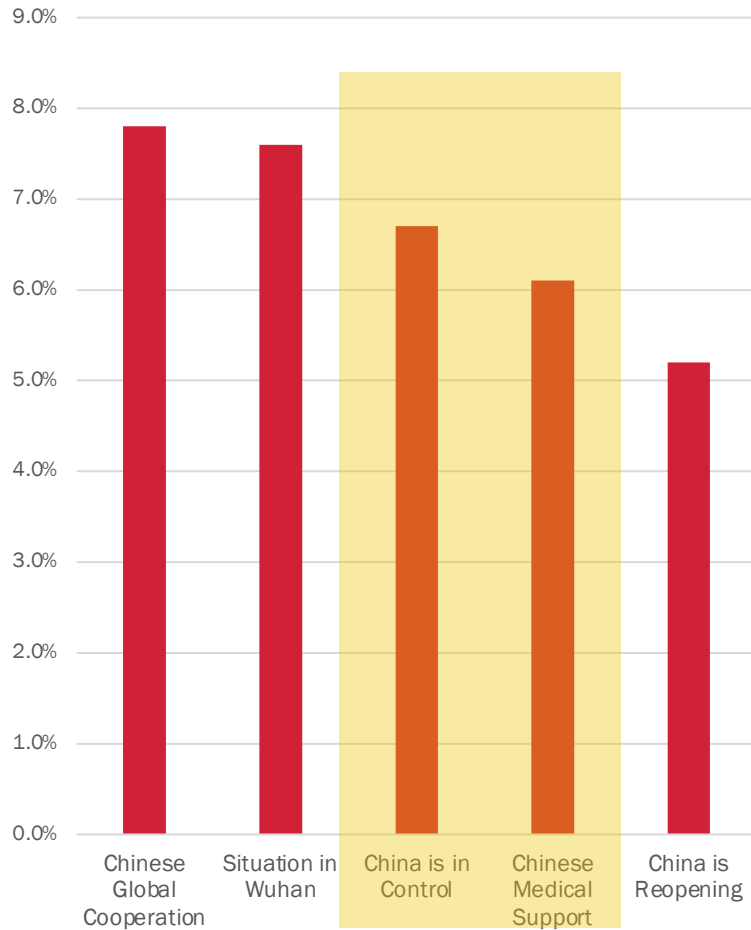
publish_date	source	country_iso	full_text	web_url	data_source	id	language_gdelt	count_china	co
2020-04-07	abidjan.net	CI	Société Hamed Bakayoko déclaré positif au coronavirus...	https://news.abidjan.net/h/671218.html	GDELT	GDELT126371	FR	0	
2020-04-27	abidjan.net	CI	Société COVID 19: Un don en vivres et non vivres offre...	https://news.abidjan.net/h/672003.html	GDELT	GDELT126521	FR	0	
2020-04-20	abidjan.net	CI	Économie Covid-19: le directeur général de IOMS peu...	https://news.abidjan.net/h/671694.html	GDELT	GDELT126580	FR	0	
2020-04-22	abidjan.net	CI	Économie Hausse du coût de la viande : les acteurs in...	https://news.abidjan.net/h/671803.html	GDELT	GDELT126626	FR	0	
2020-04-29	abidjan.net	CI	Société Secteur du tourisme et des loisirs : les résulta...	https://news.abidjan.net/h/672090.html	GDELT	GDELT126639	FR	0	
2020-03-21	abidjan.net	CI	Économie Covid-19: Les mesures urgentes prises par ...	https://news.abidjan.net/h/670625.html	GDELT	GDELT126836	FR	0	
2020-03-23	abidjan.net	CI	Société Journée mondiale de l'eau 2020 toujours d'ac...	https://news.abidjan.net/h/670673.html	GDELT	GDELT126859	FR	2	
2020-03-27	abidjan.net	CI	International New York, nouvel épicerie de la pandé...	https://news.abidjan.net/h/670803.html	GDELT	GDELT126863	FR	0	
2020-05-06	abidjan.net	CI	Société Ramadan : Hamed Bakayoko offre des vivres a...	https://news.abidjan.net/h/672361.html	GDELT	GDELT127053	FR	0	
2020-05-12	abidjan.net	CI	Société Coronavirus : Le Premier Ministre met la minis...	https://news.abidjan.net/h/672617.html	GDELT	GDELT127165	FR	0	
2020-05-14	abidjan.net	CI	Société Reprise des travaux de la Nationale A3 à Niak...	https://news.abidjan.net/h/672735.html	GDELT	GDELT127209	FR	0	
2020-05-27	abidjan.net	CI	Politique Présidentielle 2020: la révision de la liste éle...	https://news.abidjan.net/h/673528.html	GDELT	GDELT127343	FR	0	
2020-02-27	abidjan.net	CI	■	https://news.abidjan.net/h/669714.html	GDELT	GDELT127411	FR	1	
2020-04-03	abidjan.net	CI	Économie Covid- 19 Des producteurs de café-cacao r...	https://news.abidjan.net/h/671048.html	GDELT	GDELT127456	FR	0	
2020-04-05	abidjan.net	CI	Société Le préfet de Korhogo appelle la population à s...	https://news.abidjan.net/h/671111.html	GDELT	GDELT127484	FR	0	
2020-04-10	abidjan.net	CI	Région COVID-19 : Agboville lance la phase répressiv...	https://news.abidjan.net/h/671331.html	GDELT	GDELT127504	FR	0	
2020-04-12	abidjan.net	CI	Société Pâques 2020 : L'Église méthodiste plaide pour ...	https://news.abidjan.net/h/671409.html	GDELT	GDELT127515	FR	0	
2020-04-15	abidjan.net	CI	Art et Culture Covid 19: Des artistes et journalistes se...	https://news.abidjan.net/h/671480.html	GDELT	GDELT127531	FR	0	
2020-04-10	abidjan.net	CI	Société Coronavirus : Le centre de dépistage de Yopo...	https://news.abidjan.net/h/671348.html	GDELT	GDELT127533	FR	0	
2020-04-04	abidjan.net	CI	Société Covid-19 : Sensibilisation à outrance à lintéri...	https://news.abidjan.net/h/671076.html	GDELT	GDELT127552	FR	0	
2020-04-09	abidjan.net	CI	Société Covid-19: Voici les techniques utilisées par d...	https://news.abidjan.net/h/671283.html	GDELT	GDELT127592	FR	0	
2020-04-03	abidjan.net	CI	Société Treichville/ Lutte contre le Coronavirus : le co...	https://news.abidjan.net/h/671067.html	GDELT	GDELT127637	FR	0	
2020-04-15	abidjan.net	CI	Société Covid-19: L'état de Côte d'Ivoire proroge le dél...	https://news.abidjan.net/h/671500.html	GDELT	GDELT127675	FR	0	
2020-04-19	abidjan.net	CI	Société Bruno Koné fait des dons aux ressortissants d...	https://news.abidjan.net/h/671672.html	GDELT	GDELT127752	FR	0	
2020-03-06	abidjan.net	CI	Société Social: Report du gala de la Fondation Childre...	https://news.abidjan.net/h/670091.html	GDELT	GDELT128005	FR	0	
2020-03-18	abidjan.net	CI	Société Coronavirus: Latelier de formation des format...	https://news.abidjan.net/h/670503.html	GDELT	GDELT128032	FR	0	
2020-03-19	abidjan.net	CI	Politique Covid-19: La plateforme pro-Gbagbo, EDS, l...	https://news.abidjan.net/h/670561.html	GDELT	GDELT128060	FR	0	
2020-03-17	abidjan.net	CI	Société Crise du Coronavirus : Les Abidjanais vident l...	https://news.abidjan.net/h/670484.html	GDELT	GDELT128104	FR	0	
2020-03-27	abidjan.net	CI	Économie Impacts du Covid-19 dans l'UEMOA : Avis r...	https://news.abidjan.net/h/670824.html	GDELT	GDELT128219	FR	0	
2020-03-26	abidjan.net	CI	Économie Covid-19/Abidjan : Le marché gouro ferme...	https://news.abidjan.net/h/670770.html	GDELT	GDELT128249	FR	0	
2020-05-06	abidjan.net	CI	Société Lutte contre le Covid-19: démarrage officiel d...	https://news.abidjan.net/h/672345.html	GDELT	GDELT128331	FR	0	
2020-05-04	abidjan.net	CI	Société Ramadan : 400 familles démunies soulagées à...	https://news.abidjan.net/h/672295.html	GDELT	GDELT128374	FR	0	
2020-05-10	abidjan.net	CI	Société Lutte contre le Covid-19: les commerçants du...	https://news.abidjan.net/h/672540.html	GDELT	GDELT128377	FR	0	
2020-05-13	abidjan.net	CI	Société Des enseignants musulmans offrent des kits a...	https://news.abidjan.net/h/672674.html	GDELT	GDELT128427	FR	0	
2020-05-12	abidjan.net	CI	Société Covid-19: des équipements sanitaires offerts ...	https://news.abidjan.net/h/672628.html	GDELT	GDELT128430	FR	0	
2020-05-20	abidjan.net	CI	Société Nuit du Qadr: Les musulmans ivoiriens invités...	https://news.abidjan.net/h/673090.html	GDELT	GDELT128470	FR	0	

RQ₁: Amount Coverage about China & COVID

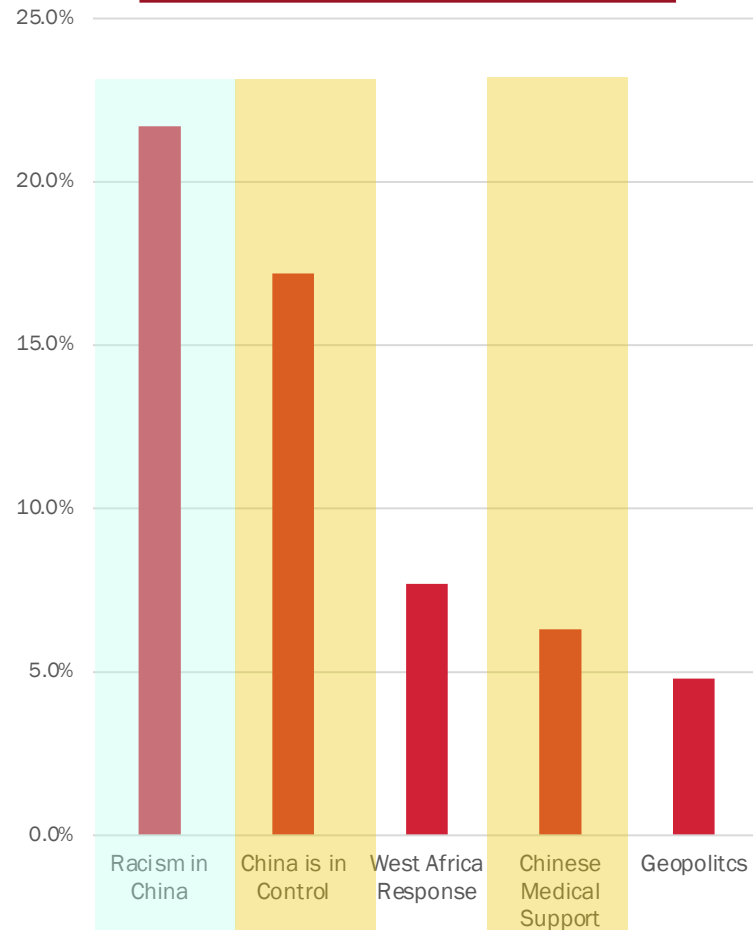


RQ₄: Narrative overlap

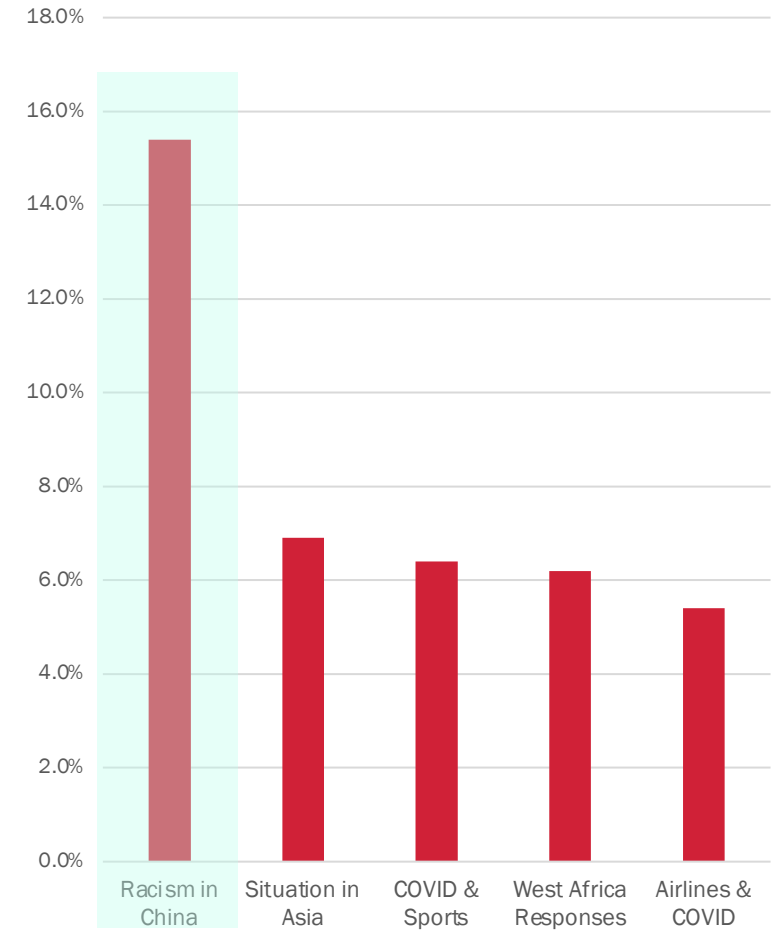
Chinese media



Gambian media



Kenyan Media



Next steps

- There are limitations on the type of questions we can answer using this approach
 - Good at **classifying documents**, not so good at making sense of **meanings**.
- Complement the results of this approach with
 - more **interviews with journalists** to understand the reasons for using/not using content;
 - additional **survey research** to understand decoding of these media texts by media consumers.

More information

- (Some of the) data and the code for this study is available in this repository:
 - https://github.com/danimadrid/africa_china_covid19
- There's more details on the process that I describe here in here:
 - <https://doi.org/10.1080/23743670.2020.1820885>

Dani Madrid-Morales, PhD
dmmorales2@uh.edu
[@DMadrid_M](#)
www.danimadrid.net