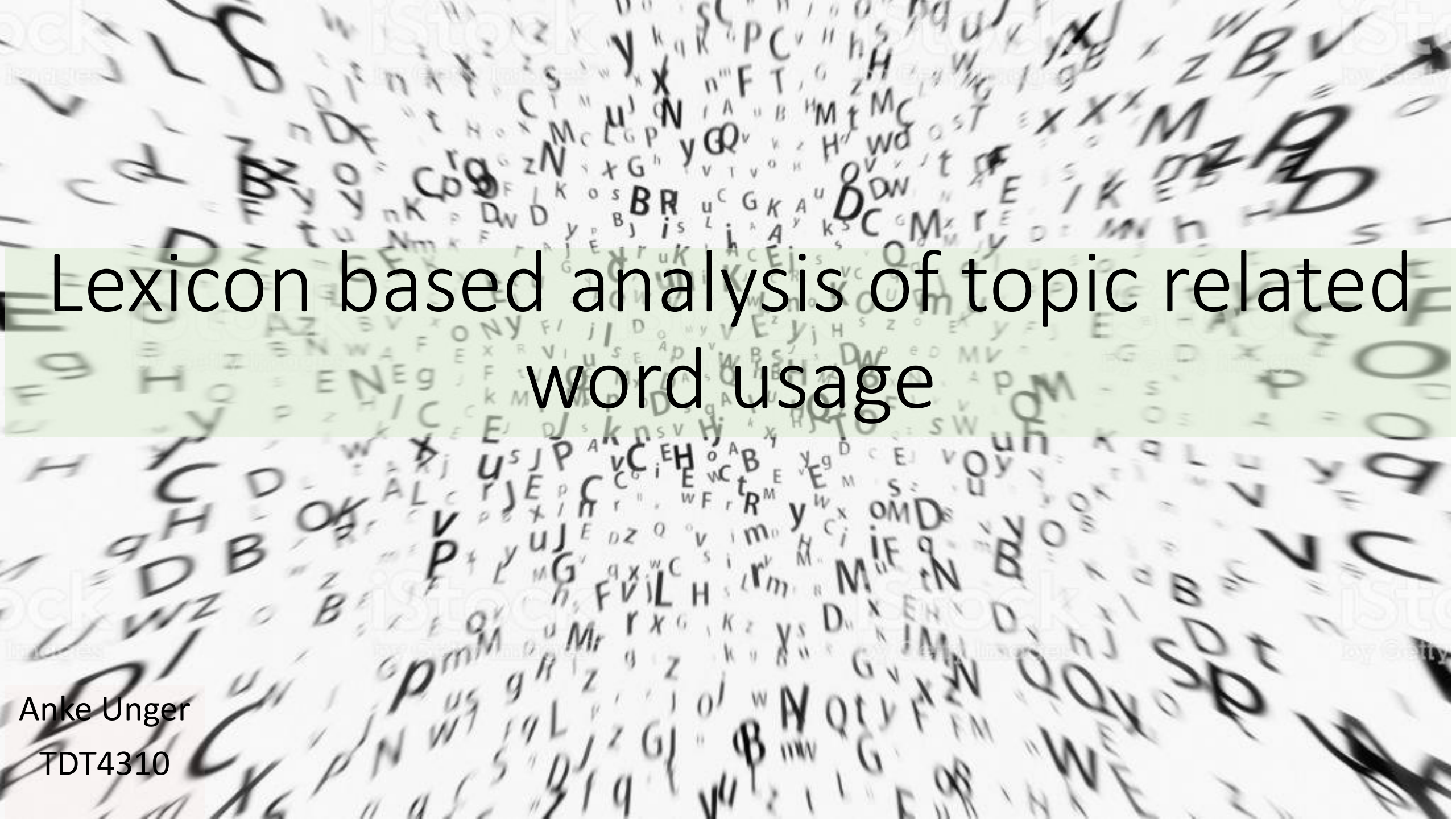


Lexicon based analysis of topic related word usage

Anke Unger

TDT4310

Overview

WHY is the analysis interesting?

HOW is it build?

WHAT are the results?

WHERE are the fallbackpoints?

WHAT can we take away?

How are certain topics depicted in the media?

"All the News That's Fit to Print"

The New York Times

CITY EDITION

VOL. CXVIII ..... No. 40,721 — NEW YORK, MONDAY, JULY 21, 1969 — 10 CENTS

MAN WALKS ON MOON

ASTRONAUTS LAND ON A PLAIN AFTER STEERING PAST CRATER

Voice From Moon: 'Eagle Has Landed'

NEIL A. ARMSTRONG

EDWIN E. ALDRIN JR.

Bleak, Rocky World Seen From Module

By JOHN NOBLE WILFORD

VOYAGE TO THE MOON

By Anthony MacLEISH

Boulders May Prove Scientific Boons

By WALTER SULLIVAN

Luna Dips to 10 Miles From Moon; Surmises Soar

Source: https://www.chroniclelive.co.uk/news/north-east-news/how-sunday-sun-first-newspaper-16561422

The Times Dispatch

WHOLE NUMBER 18357. RICHMOND, VA., TUESDAY, APRIL 16, 1912. PRICE TWO CENTS

Titanic, Giant White Star Liner, Sinks After Collision With Iceberg on Her Maiden Voyage, and 1,800 Lives Are Reported Lost in World's Greatest Marine Disaster

WIRELESS CALLS SEND VESSELS RUSHING TO AID OF SEA COLOSSUS

Through the Night They Drive With Full Speed to Reach Titanic.

MEANTIME, WORLD WAITS IN AGONY OF SUSPENSE

OF ALL ON BOARD ONLY 675 KNOWN TO HAVE ESCAPED DEATH IN OCEAN

Those Rescued Mostly Women and Children, Who Were Taken Off in Boats.

BITS OF WRECKAGE ALL THAT IS LEFT OF GREAT VESSEL

THE TITANIC

COL. JOHN JACOB ASTOR IS AMONG DROWNED

HIS WIFE IS SAVED

CARRIED NOTABLE PASSENGER LIST

FINANCIAL BLOW IS VERY SEVERE

HAD $5,000,000 INSURANCE

OFFICIALS CONCEDE GREAT LOSS OF LIFE

PEOPLE ON SHIPS GET FIRST NEWS

STILL HOPING FOR BEST

PREVIOUS BIG SEA DISASTERS

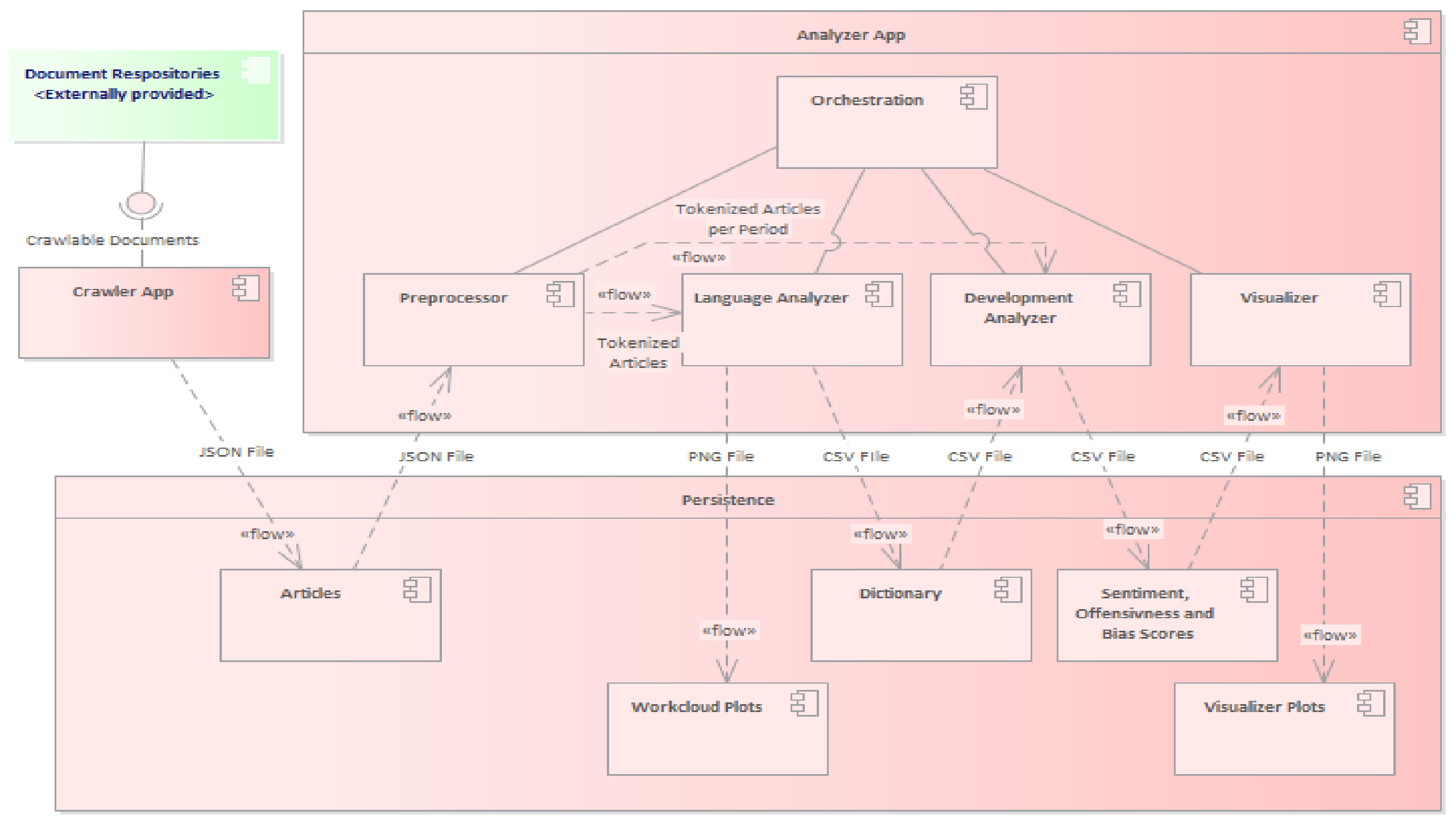
Source: https://www.ultimatetitanic.com/titanic-news/?category=Sinking

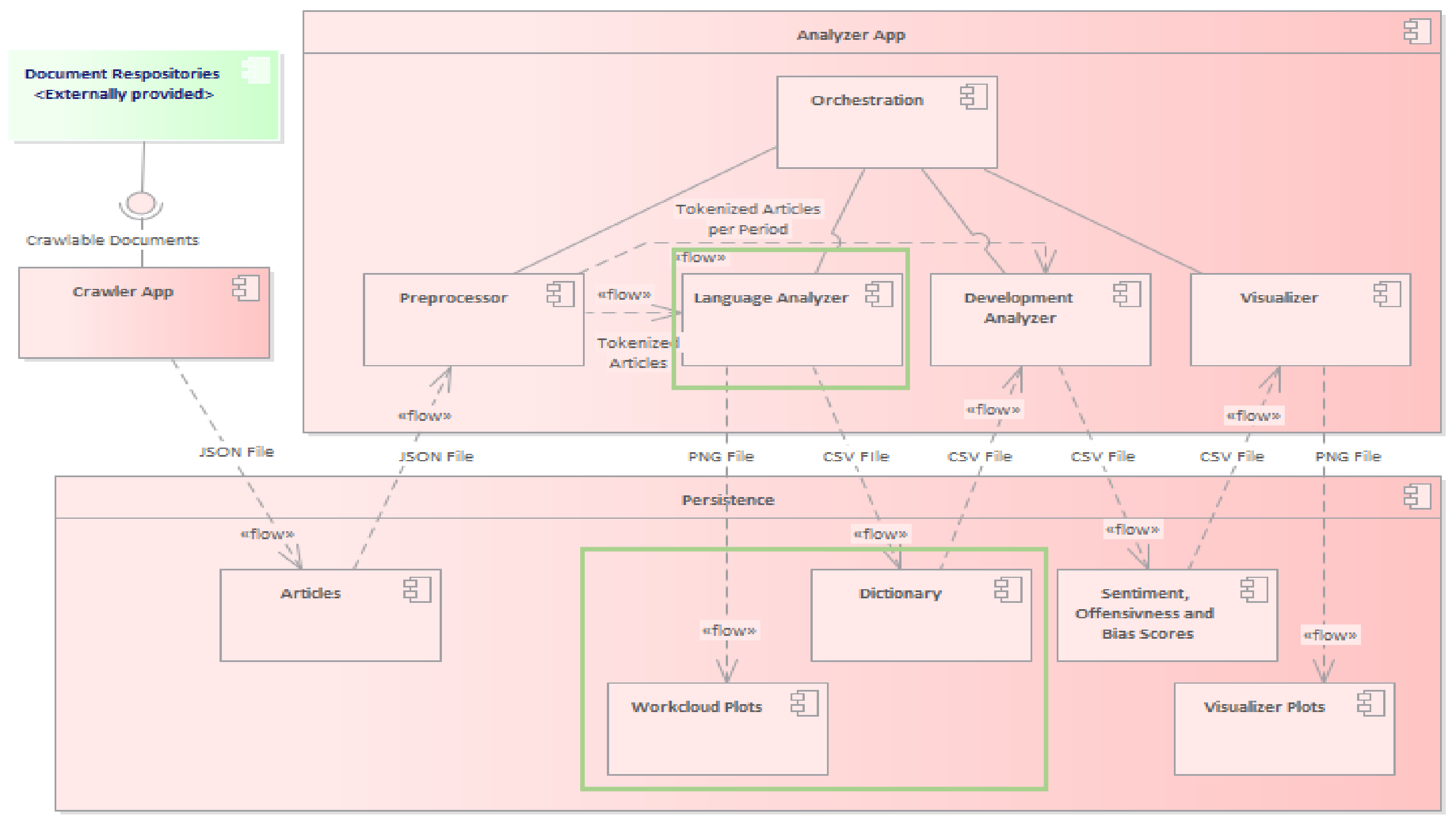
Supervised VS Semi-Supervised training

- Language is flexible
- Most tagged sets for short texts
  → Narrow applicability

Idea:

- Lexicon based for biased or offensive language lexica
- Semi-Supervised for sentiment lexicon
  → Test on articles regarding LGBT-community and Corona Crisis


Document Respositories
<Externally provided>
Crawlable Documents
Crawler App
Analyzer App
Orchestration
Tokenized Articles
per Period
«flow»
Preprocessor
«flow»
Language Analyzer
Development
Analyzer
Visualizer
Tokenized
Articles
«flow»
«flow»
«flow»
JSON File
JSON File
PNG File
CSV File
CSV File
CSV File
CSV File
PNG File
Persistence
«flow»
«flow»
«flow»
Articles
Dictionary
Sentiment,
Offensivness and
Bias Scores
«flow»
«flow»
Workcloud Plots
Visualizer Plots


Document Respositories
<Externally provided>
Crawlable Documents
Crawler App
Analyzer App
Orchestration
Tokenized Articles
per Period
«flow»
Preprocessor
«flow»
Language Analyzer
Development
Analyzer
Visualizer
Tokenized
Articles
«flow»
«flow»
«flow»
«flow»
JSON File
JSON File
PNG File
CSV File
CSV File
CSV File
CSV File
PNG File
Persistence
«flow»
«flow»
«flow»
«flow»
«flow»
Articles
Dictionary
Sentiment,
Offensivness and
Bias Scores
Workcloud Plots
Visualizer Plots

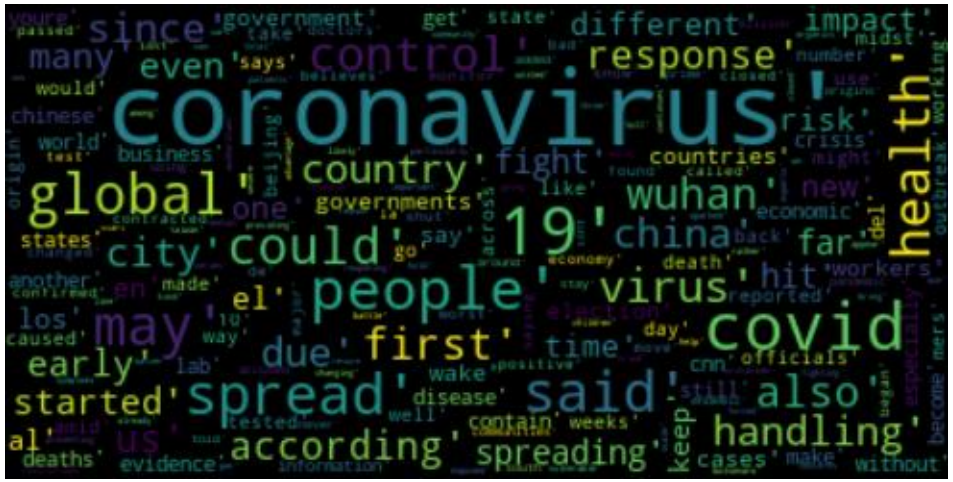
Crawling and Preprocessing

[covid, cov, corona, sars, outbreak, pandemic, virus]

[…, the, war, of, words, between, the, trump, administration, and, china, over, the, origin, of, coronavirus, has, escalated, as, president, trump, is, going, out, of, his, way, to, call, it, the, chinese, virus, and, china, has, embarked, on, a, propaganda, campaign, to, shift, blame, for, the, pandemic, away, from, the, chinese, communist, party, handling, of, the, crisis, …]

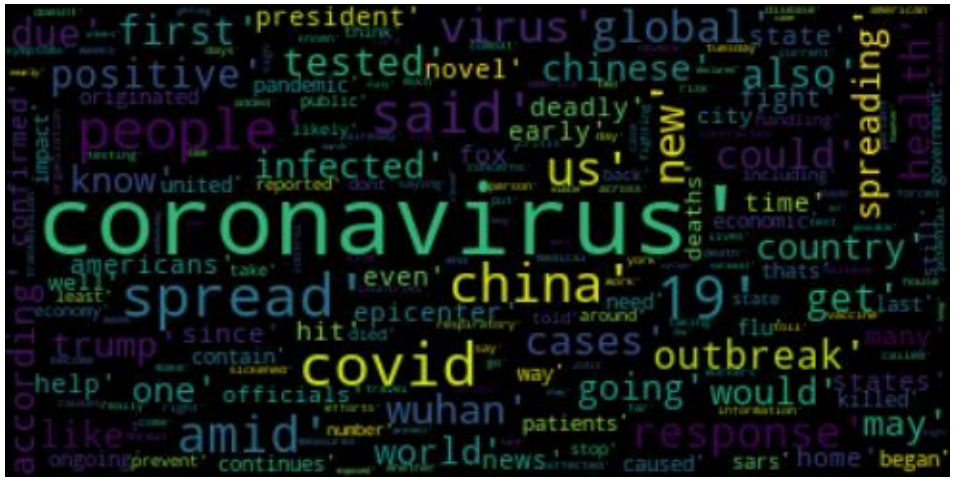
from https://www.foxnews.com/politics/gloves-off-over-coronavirus-trump-administration-china-escalate-clash-over-origins

Wordclouds Corona Crisis and LGBT



Corona-context CNN

LGBT related articles 1986-1999



Corona-context Fox News

LGBT related articles 2000-2015

Sentimentlexicon extension

- Conjunctions identified

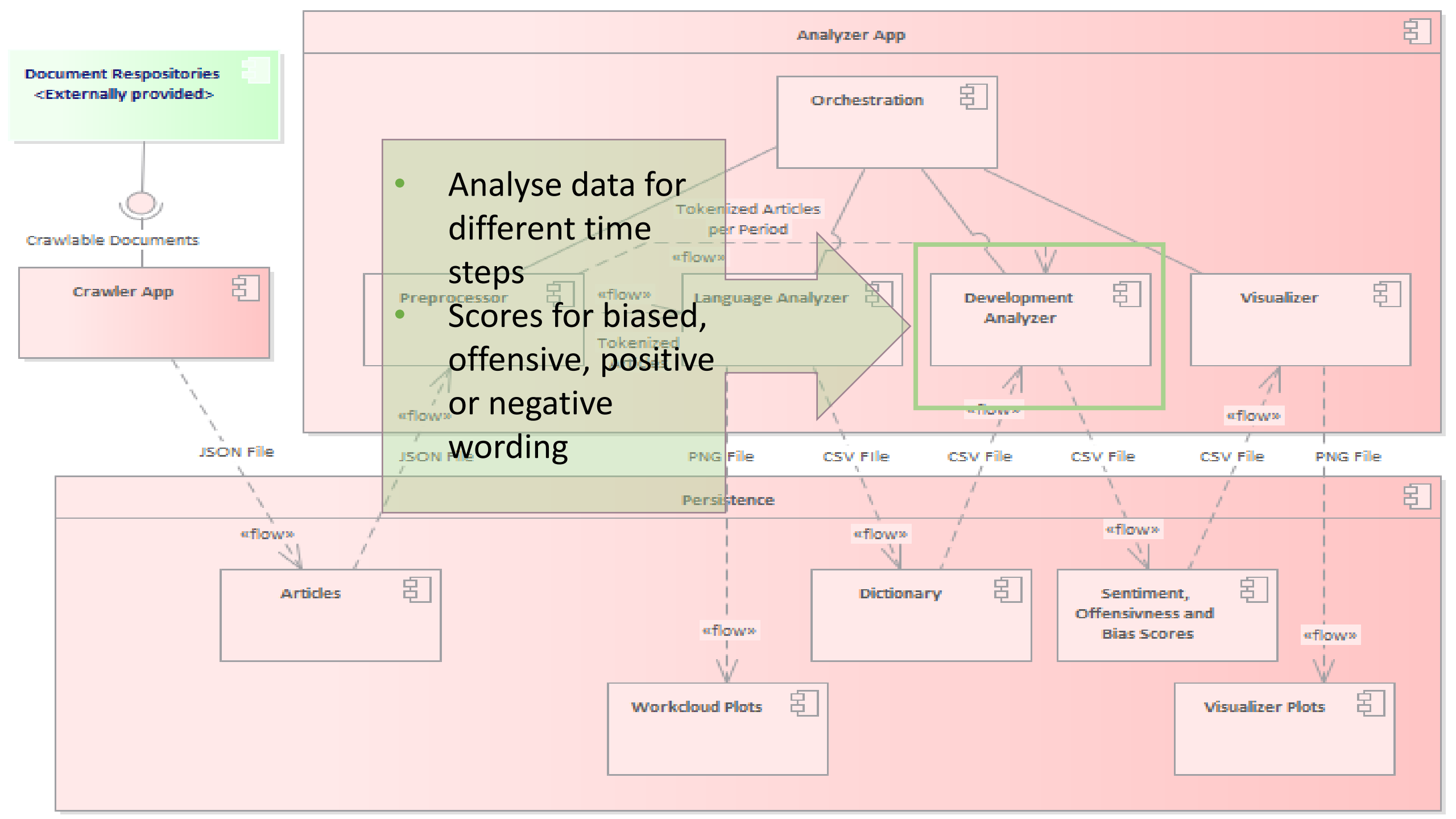
  ‘but’ → opposite polarity

  ‘and’ → same polarity

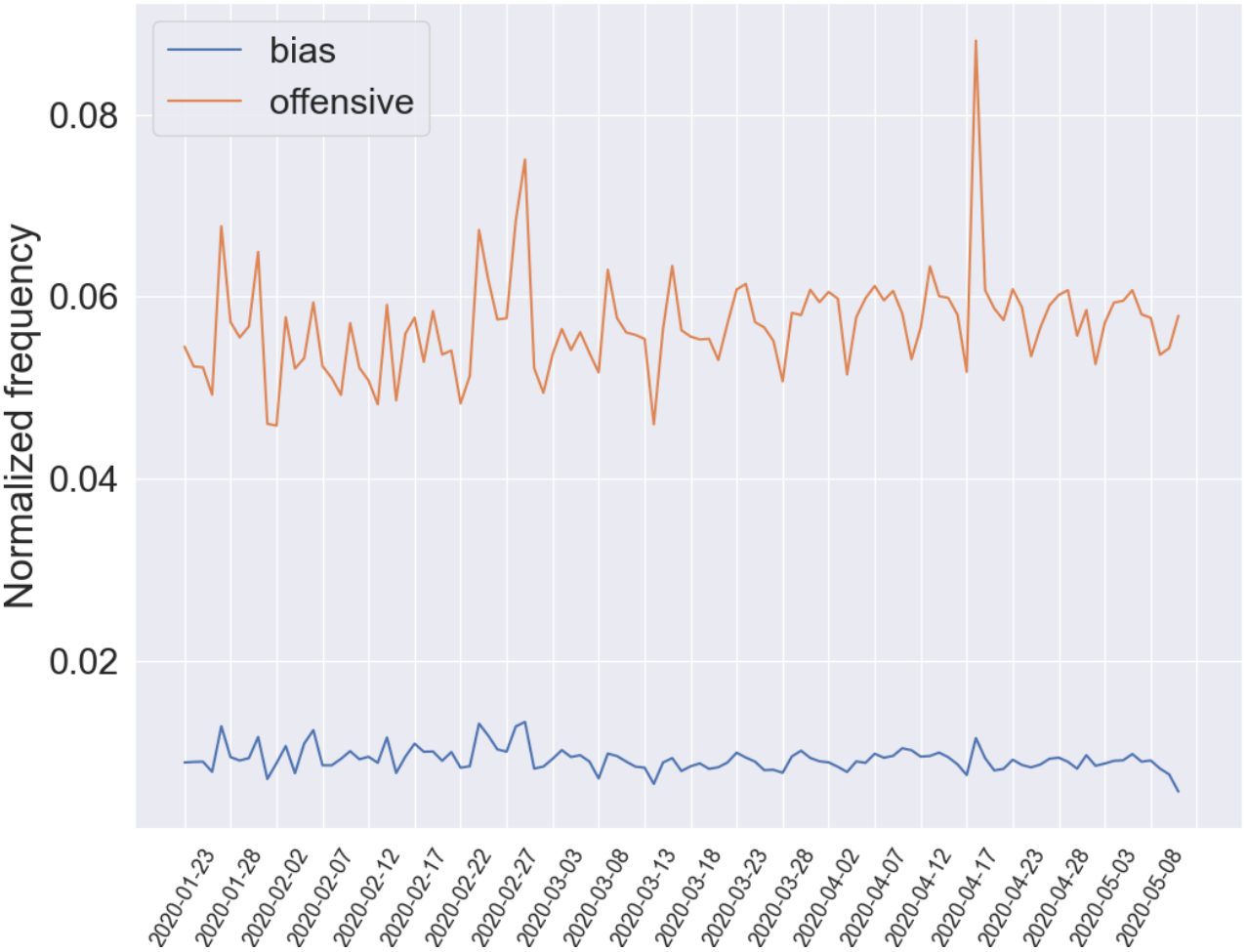
- Weighted with Inverse Document Frequency (IDF)

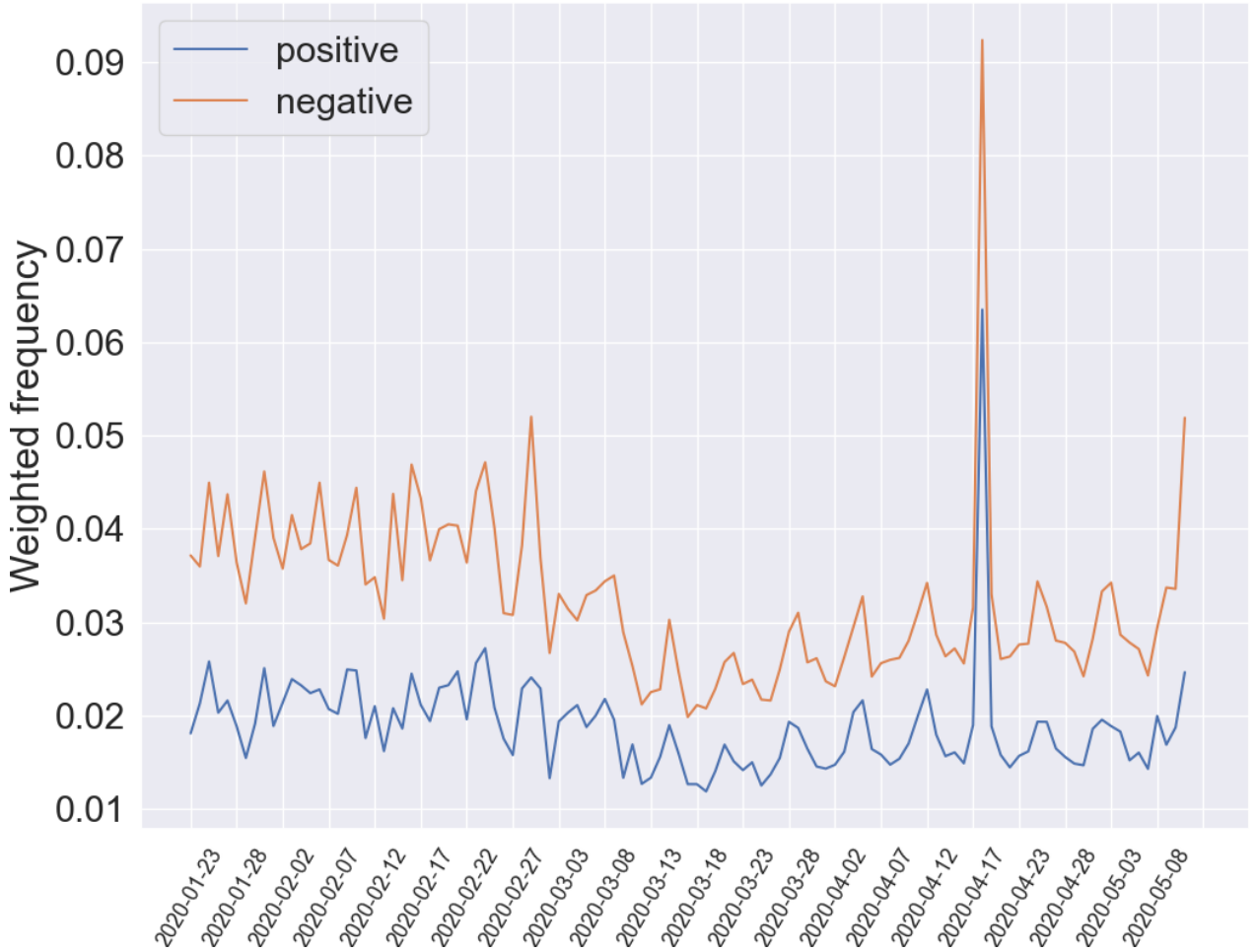
Result:

’economic’, ’responsible’, ’stable’, ’angerthe, ’heat-sensitive’

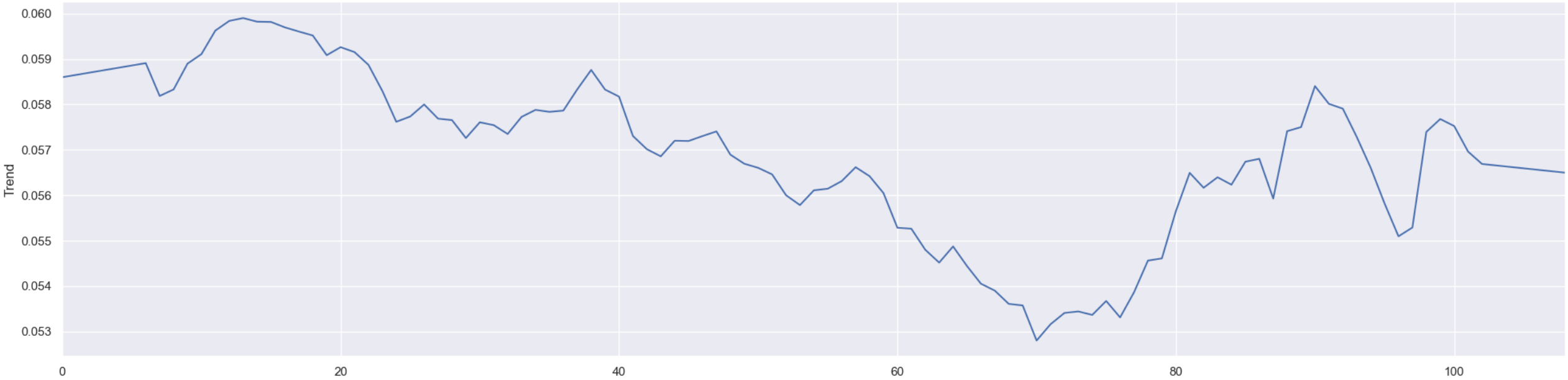

Document Respositories
<Externally provided>
Crawlable Documents
Crawler App
Analyzer App
Orchestration
Tokenized Articles
per Period
«flow»
Preprocessor
«flow»
Language Analyzer
Tokenized
Development
Analyzer
Visualizer
«flow»
«flow»
«flow»
JSON File
JSON File
PNG File
CSV File
CSV File
CSV File
CSV File
PNG File
Persistence
«flow»
«flow»
«flow»
«flow»
«flow»
Articles
Dictionary
Sentiment,
Offensivness and
Bias Scores
Workcloud Plots
Visualizer Plots
Analyse data for different time steps
Scores for biased, offensive, positive or negative wording

Development over time





Trend?



Trend in usage of possibly offensive language

Take home

- Lexicon based approaches demand less previous knowledge
- Frequency of word usage helps to determine overall topics of articles
- Frequency of word usage helps to give suggestions how controversial a topic is

https://www.shutterstock.com/de/search/take-home