Computational Analysis of Mass Media

2014 © Mediator Project

http://mediaanalysistools.github.io

Project motivation

Mass media landscape was transformed by Internet:
- real-time access to ever-growing amounts of multimedia information
- anyone can be a news producer

but:
- quick distribution of false or unsubstantiated claims and biased news stories
- easier and imperceptible propagation of ideas by paid or ideologically driven groups of people

There is **rich literature** on theoretical and experimental studies of propaganda, media bias, and activity in social networks, but significantly **less computational tools** for media analysis available for the community.

Missinformation

Mass media are known for spreading factually incorrect information:

- **misquotation** (e.g., the British Homeopathic Association was criticized for providing out-of-context quotations from scientific papers as evidence ▶)

- **misuse of** factual, statistical, or historical **data** (e.g., a factually incorrect claim by the Texas Governor Rick Perry about crimes committed by illegal aliens ▶)

- **fake or irrelevant pictures** illustrating news (e.g., a Syrian child sleeping in between his parents' graves ▶)

Media bias

- the bias of news producers in the selection of events and stories that are reported and how they are covered in terms of linguistic and visual means.

- is inherent to the news production process
- for news readers, it is difficult to analyze mass media critically and detect the bias
- even for news producers, the bias in their own writing is often not obvious (cf. _Media Lens_)
- Selective exposure theory: people tend to seek out information confirming their beliefs and avoid challenging information
- online information customization multiplies this effect
- result: opinion polarization and misunderstanding of critical issues

Manufacted opinion spread

- shift from traditional media (newspapers and television) to social media

- every individual is a potential news producer

- Influencers theory: news readers trust other "regular people" who share their interests

- Astroturfers – paid or ideologically driven people massively producing posts, tweets, and comments for promoting a certain product or viewpoint

Project goals

Provide media scientists, journalists, news producers, and news readers with computational tools for

1. **Tracking distribution** of a media item (text, image), i.e. looking for information sources, citations, modifications.
2. **Finding contradictions and support** for news statements.
3. **Detecting media bias** and corresponding interest groups.
4. **Detecting manufactured opinion spread** (astroturfing).

As a result, build an **interactive graphical Internet news service** performing a multi-faceted analysis of media in multiple languages.

State of the art: manual media analysis

- The Glasgow University Media Project

- *FAIR* media watch group: criticism of media bias and censorship

- *Media Lens*: bias in the news articles

- *FactCheck* and *Politifact*: monitor public speeches and check the factual accuracy of what is said by U.S.
  politicians

- *Project Censored*: tracking published news and selecting news stories of social signicance that have been overlooked, under-reported, or self-censored

- *SourceWatch.org*: manually curated information about the reputation and known biases of various sources

- …

State of the art: computational tools

- Ranking originality
  - *Google News*
- Validating web content
  - *Dispute Finder*:   collects disputed claims
  - *WikiTrust*: online Wikipedia reputation system
- Balanced reading guide
  - *Balancer*: shows users feedback about the political lean of their reading
  - *NewsCube*: classifies news articles according to viewpoints and provides multiple viewpoints to news readers
- Meme analysis
  - *NIFTY*: meme tracking
  - *Truthy*: astroturfing detection for memes

Proposed solutions

Tracking media items

- Access news articles and supporting information (Wikipedia, statistics, etc.) through RSS

- Given a query (link, text, image), find documents containing similar media items using a search API

- Extract publication date, author, source

- Detect text and image modifications

http://mediaanalysistools.github.io/prototype.html

Finding contradicting and supporting claims

- Extract narratives from text (atomic events that share event participants)
- Align narratives in related documents
- Classify aligned narratives as supporting or contradicting

Iran has been working on a nuclear weapon

contradiction → *Iran is not building a nuclear weapon*

support → *Iran has at least one nuclear bomb*

- Detect multimedia inconsistencies (image captions vs article content)

Media bias types

Two types of media bias

- **selection** bias - the choices of what aspects of events to cover

- **presentation** bias - the choices of linguistic and visual means used to describe the events

Selection bias example

Thousands of South Korean police and soldiers on Thursday evicted the remaining occupants and anti-U.S. civic activists from an area designated for expanded U.S. military facilities south of Seoul...

Another violent showdown looms as groups protesting against the relocation of U.S. military bases to Pyeongtaek, Gyeonggi Province are gearing up for a massive 10,000-strong protest this weekend...
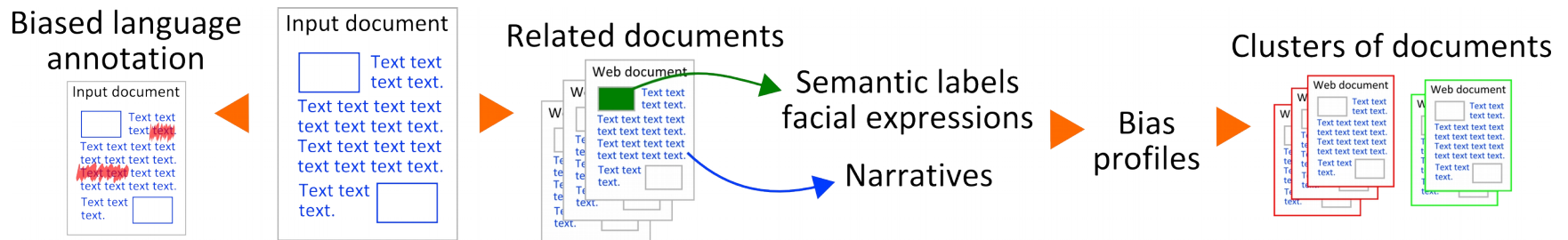
Presentation bias examples

The inadvertent killing by Israeli forces of Palestinian civilians usually in the course of shooting at Palestinian terrorists...

...Palestinian public protests and civilian demonstrations were answered brutally by Israel, which killed tens of unarmed protesters.

Media bias detection

- Generate bias profiles
  - atomic events described in text + representation means
  - atomic events recognized in images + facial expression recognition
- Classify related documents according to bias profiles, detect interest groups
- Label biased language

Biased language annotation

Input document
Text text text text.
Text text text text text text text text.
Text text text text text text text text.
Text text text.

Input document
Text text text text.
Text text text text text text text text.
Text text text text text text text text.
Text text text.

Related documents

Web document
Text text text text.
Text text text text text text text text. text text text text.
Text text text.

Semantic labels facial expressions

Narratives

Bias profiles

Clusters of documents

Web document
Text text text text.
Text text text text text text text text. text text text text.
Text text text.

Web document
Text text text text.
Text text text text text text text text. text text text text.
Text text text.

Detecting manufactured opinion spread

Goal: checking if an input media item (text fragment, hashtag, url, image) is a part of a coordinated campaign in social media.

- **Astroturfing pattern**: At the beginning, a large number of messages with a similar content produced by a small number of users. Later on, many of the users involved may be legitimate users.

- **Classify media items** in the input document as **"truthy"** (falsely propagated) or **"legitimate"** using features like:

    - # unique users posting the media item

    - temporal distribution of the media item

    - statistics about users propagating the media item

    - # reposts/retweets vs. injection points

    - ...

Study of user behavior

Experimental evaluation of each media analysis component:

- functionality and usability

    - ask users to perform small controlled tasks (e.g., *Find all sources of documents contradicting the given document*, *Find the first publication of the given media item*)

    - measure time users needed to conduct a task

    - measure the amount of errors

Study of user behavior

- effectiveness

Preprocessing:

- select a couple of events recently covered in the news

- extract two articles per event presenting diverse views

- perform full computational analysis of the articles

Experiment:

- ask subjects about their position towards the selected events

- let subjects explore the articles in the news service

- ask subjects about credibility of the article claims and their perspective

Evaluation:

- clickstream analysis

- compare the viewpoints of the subjects before and after using the media analysis tools
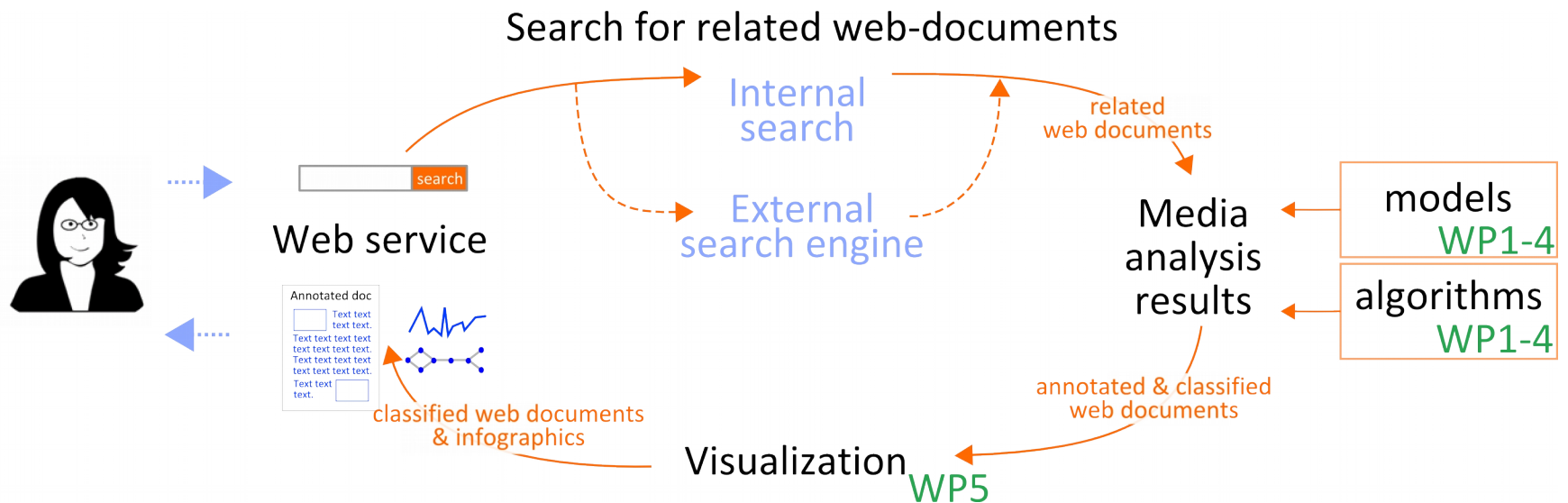
We need feedback and input from
media scientists, journalists,
news readers:

What is useful for you?