



Identification of Media Bias



- Automated identification of media bias is an applied NLP research.
- The project aims identifying text snippets, visual elements, or ideas conveyed in the texts that make news consumers perceive information in a biased way, i.e., word choice and labeling that portrayed politicians in a specific way.
- The main goal is to identify instances of bias in multiple forms and increase bias awareness of the news readers.
- Media bias goes hand-in-hand with fake news detection but focuses mostly on the official news outlets.
- Media bias detection is an interdisciplinary project between computer science, political science, digital humanities, psychology, computational linguistics.
- The project covers NLP areas such as:
 - Coreference resolution
 - Sentiment analysis
 - Hate speech detection
 - Dependency parsing
 - Information extraction (e.g., named entity recognition)

A complete list of Media Bias topics visit our [website!](#)



How to resolve "invade the country" = "cross the border"?!

Background

When reporting about events, journalists use different words to describe the same actors and entities, often based on personal or the outlet's political or ideological views. News consumers are highly influenced by a non-objective reporting style. Current state-of-the-art cross-document co-reference resolution (CDCR) systems still lack robust approaches to resolve mentions of high lexical diversity and complex non-named-entity concepts.

Goal

Develop a novel CDCR model based on Transformers by using transfer learning from paraphrase identification to identify coreferential mentions of high lexical diversity.

Tasks

- Review literature about CDCR;
- Review literature on Transformers, transfer learning, and datasets for CDCR and paraphrase identification;
- Design and train a neural network model to resolve noun and verb phrases referring to the same concepts;
- Evaluate the algorithm in a diverse collection of CDCR datasets.

CNN	Al Jazeera
UK soldiers cleared in Iraqi death	British murderers in Iraq acquitted
Seven British soldiers were acquitted on Thursday of charges of beating an innocent Iraqi teenager to death with rifle butts.	The judge on Thursday dismissed murder charges against seven soldiers, who are accused of murdering Iraqi teenager.

Anastasia Zhukova
zhukova@gipplab.org



Felix Hamborg
hamborg@gipplab.org



Background

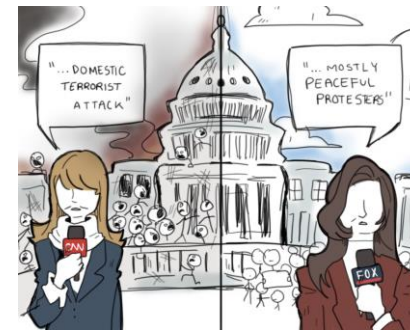
When reporting about events, journalists use different words to describe the same actors and entities, often based on personal or the outlet's political or ideological views. Identity coreference relations typically link mentions such as "Donald Trump" and "the president" that report about the true facts. On the contrary, loose coreference or bridging relations may convey bias, e.g., in "Donald Trump's 'Impulsive' Decision-making" a metonymy relation between "Trump" and "decision-making" frames Donald Trump as an impulsive person. Bias by word choice and labeling yield a non-objective reporting style and influences news consumers.

Goal

Create a new CDCR dataset by researching how phrases become coreferential with relations of varying strength, how these relations facilitate bias by word choice and labeling, and influence on news readers' perception of the information.

Tasks

- Review literature review about 1) types of coreference, bridging, and near-identity relations, 2) which types of relations do the datasets for (cross-document) coreference resolution include or excluded from annotations.
- Explore how coreference and bridging relations influence in news readers' reasoning.
- Create a coding book to annotate coreferential mentions and their relations to identify cases of bias by word choice and labeling.
- Annotate a new CDCR dataset that focuses on high lexical diversity.



Anastasia Zhukova
zhukova@gipplab.org



Felix Hamborg
hamborg@gipplab.org

