

# Creation of a Machine Learning Media Bias Dataset

## Background

This project will prepare machine learning for a large, international study that aims on collecting annotated media bias in news texts to be applied for machine learning. We intend to use human-identified media bias, based on word choice and labelling, to filter out underlying concepts of media bias, to evaluate existing approaches, and to detect media bias ourselves, using the whole structure of annotations to train a machine on identifying the same concepts.

## Goal

Your task will be to perform such a study on a smaller level.

## Tasks

- Get familiar with media bias & survey methodology
- Perform a pre study for a later, bigger experiment: Develop an easily understandable annotation survey which makes people annotate bias. The participants will be shown different texts or text excerpts and will then mark what they believe is media bias. Communication is a key task in this project.
- Try out different machine learning technologies to see which could achieve the highest performance for such a task.
- Analyze the results of your pretests. Evaluate, repeat and Re Evaluate.



Timo Spinde

Timo.Spinde@uni-konstanz.de



Felix Hamborg

Felix.Hamborg@uni-konstanz.de



# Creation of a morphological bias resource for the German language

## Background

The German word Flüchtlinge is one example for a morphologically degrading entity. Its general impression is influenced by the derivational component -ling, which frequently is and has been origin for derogatory replacement constructions, e.g. Sonderling, Schönling or Schwächling. In this bachelor thesis project, you will work on a – at one future point – very resource independent measure of (media) bias at the sub-word level. To achieve this, you will study morphological patterns in different lexica, manually and automatically. You will then bring the patterns into the context of media bias.

## Goal

Use machine learning and NLP to compare sentiment and morphological patterns in lexica.

## Tasks

- Identify key sentiment and polarity analysis resources
- Reduce their words to prefixes, stems and suffixes and map these to polarity to identify patterns.
- Identify the relation between “polarity” and “bias”
- To identify the patterns, use e.g. machine learning / network analysis / statistical modelling.
- Compare your results to the words used by different newspaper outlets within several small case studies



Timo Spinde

Timo.Spinde@uni-konstanz.de



Tatjana Scheffler

tatjana.scheffler@uni-potsdam.de



## Background

Do you say: "Refugees will be distributed among all of Germany"? Or do you say: "Islam invades Germany?" Media bias is one of the most trending scientific topics today, it affects us, everywhere, anytime. We developed an approach that improves performance of identifying medium-specific bias language. In the next step, this approach will become automated. For technical details and some related frameworks, please see <https://bit.ly/2nwHiFU>.

## Goal

Automate a method that creates bias lexica from specific sources.

## Tasks

- Get familiar with media bias & semantic models.
- Scrape an English data set.
- Build a robust automated or semi-automated algorithm to detect media bias based on biased inputs.
- Test and evaluate the approach.



Timo Spinde

Timo.Spinde@uni-konstanz.de



Felix Hamborg

Felix.Hamborg@uni-konstanz.de

