

# Media Bias in German Online Newspapers

Alexander Dallmann  
Data Mining and Information  
Retrieval Group  
University of Würzburg  
dallmann@informatik.uni-  
wuerzburg.de

Daniel Zoller  
Data Mining and Information  
Retrieval Group  
University of Würzburg  
zoller@informatik.uni-  
wuerzburg.de

Florian Lemmerich  
Computational Social  
Science Group  
GESIS - Leibniz Institute for  
the Social Sciences  
florian.lemmerich@gesis.org

Andreas Hotho  
Data Mining and Information  
Retrieval Group  
University of Würzburg & L3S  
Research Center  
hotho@informatik.uni-  
wuerzburg.de

## ABSTRACT

Online newspapers have been established as a crucial information source, at least partially replacing traditional media like television or print media. As all other media, online newspapers are potentially affected by media bias. This describes non-neutral reporting of journalists and other news producers, e.g., with respect to specific opinions or political parties. Analysis of media bias has a long tradition in political science. However, traditional techniques rely heavily on manual annotation and are thus often limited to the analysis of small sets of articles. In this paper, we investigate a dataset that covers all political and economical news from four leading German online newspapers over a timespan of four years. In order to analyze this large document set and compare the political orientation of different newspapers, we propose a variety of automatically computable measures that can indicate media bias. As a result, statistically significant differences in the reporting about specific parties can be detected between the analyzed online newspapers.

## Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

## General Terms

media analysis, computational social science

## 1. INTRODUCTION

Online newspapers have been established as a crucial information source in modern societies, at least partially replacing traditional media like television or print media. As all other

media, online newspapers are potentially affected by media bias. This describes non-neutral reporting of journalists and other news producers, e.g., with respect to specific opinions or political parties. Identifying and recognizing biases in media is crucial for an open society due to the large influence of the news media as the *fourth estate*. Additionally, making biases transparent can also have practical implications: it allows journalists and publishers to assess their own work objectively, and readers to interpret statements of that source in the correct context or choose an online newspaper according to their own political preference. Analysis of media bias has a long tradition in political science. However, many traditional techniques require access to a wide range of newspapers and rely heavily on manual annotation and are therefore often limited to the analysis of small sets of articles and thus focus on small time spans, e.g., election campaigns. This paper investigates how online newspapers that are freely available on the web can be analyzed automatically with respect to potential bias towards political parties.

In this paper, we are not interested in the overall political orientation of online news media as a whole, but focus on a comparative analysis in order to identify relative biases between online newspapers. In that direction, we focus on the German online newspaper landscape. We collected a large dataset composed of all articles published in politics and economics sections of four leading online newspaper in Germany, that is, *faz.net*, *spiegel.de*, *taz.de*, and *zeit.de*, in a four-year period.

Thus, the contribution of this paper is twofold: First, we discuss several measures that allow us to identify potential bias towards a political party in online newspapers in an automatic manner. This not only enables the analysis of large document sets with reasonable costs, but also grants results that are not influenced by the subjectivity of human annotations. Second, we show for our exemplary dataset what biases four leading German online newspapers exposed in the analyzed time frame in comparison to each other. In that regard, we can identify statistically significant differences in the coverage of different parties among the analyzed online newspapers.

The rest of the paper is structured as follows: Section 2 reviews related work. Then, Section 3 introduces a vari-

ety of metrics that indicate potential bias towards political parties in online newspapers. Next, Section 4 presents the utilized datasets. Section 5 reports on experimental results. Afterwards, Section 6 discusses limitations of the proposed approach. Finally, Section 7 concludes the paper with a summary and an outlook on future research directions.

## 2. RELATED WORK

The analysis of media bias is a well developed and active research field in political sciences, but traditionally relies heavily on manual annotation of articles. This not only makes analytical studies very costly, but also introduces another form of bias through the annotators. Therefore, “text as data”, i.e., the automated analysis of large text corpora, has been recognized as an useful tool in political science in the last years, cf. [3]. In that direction, large sets of parliamentary speeches [11], the political orientation of parties [8, 12, 9], party classification of speeches [17] and the distribution of topics in congressional bills [5] have been investigated. In contrast to these approaches, this paper focuses on the large-scale analysis of online newspapers.

Literature on media bias analysis distinguishes between three types of media bias, that is, *gatekeeping bias*, *coverage bias* and *statement bias* [1]. Gatekeeping bias describes the selection of stories out of the potential stories and is generally hard to quantify. Coverage bias expresses how much space political positions (or in our case, parties) receive in media. For traditional newspapers, this has been measured by column inches of paper covered [14], see also [1]. As an alternative, also counting occurrences in the headlines has been performed for this task [13]. While counting occurrences is a tedious task to do manually for full texts of articles, it is easy to perform in an automatic approach as advocated in this paper. Statement bias denotes how an author’s own opinion is woven within a text. Traditionally, this is often handled by manual annotation of whether a text is “favorable“ to a party or not. As one of many examples, Ho and Quinn investigated the political position of media in the US party system by labeling and analyzing the agreement of editorials with Supreme Court decisions. For their study they employed a team of 14 law students and manually labeled 1500 editorials [7]. These efforts clearly illustrate the advantages of automatic approaches as proposed in this paper. Regarding automatic analysis, Groseclose and Milyo measure bias in newspaper media by first computing a score for think tanks based on the citations of Democrat or Republican party members. Using these scores, newspapers are then evaluated indirectly by investigating which think tanks are referenced [4]. Such a method could be additionally implemented to augment the analysis in our setting. However, results could be expected to be less significant for the German political landscape since think tanks are mostly publicly funded and obligated to political neutrality. Gentzkow and Shapiro investigate overall bias of US newspapers by determining the number of occurrences of typical phrases (as identified by a quantitative analysis of congressional records) for Democratic and Republican congressman in newspapers [2]. While this is related to our method for measuring vocabulary similarity to party manifestos, unlike our work their approach is tailored to the US two-party system. Additionally, this paper uses a variety of different measures for a more comprehensive picture and is focused on online newspapers.

## 3. ANALYSIS

Measuring bias in printed media is a long-term research topic in political science. However, established methods often involve manual inspection of documents, and are difficult to apply automatically on a large corpus. In addition, existing methods are often tailored to the two-party system as it is established in the United States, but are less suited to a multi-party system as it exists, e.g., in Germany. Therefore, we present in the following a set of metrics that indicate possible bias of newspapers towards certain political parties in our corpus of German online newspapers. We start by introducing some notations.

### 3.1 Notation

The corpus for each online newspaper  $N = \{d_{N,1}, \dots, d_{N,k}\}$  consists of a set of documents (articles). For each document, the title  $T(d)$ , the full text  $F(d)$ , and a set of keywords  $K(d)$  from the HTML header of the online article are available. If a title  $T(d)$  contains a subsequence  $s$  of one or more words, then this is denoted in this paper as  $s \sqsubseteq T(d)$  (or  $s \sqsubseteq F(d)$ ,  $s \sqsubseteq K(d)$  analogously). Furthermore, the length of the respective article in number of words is denoted by  $|F(d)|$ .

We focus our analysis on a set of political parties  $\mathcal{P} = \{P_1, P_2, \dots, P_p\}$ . Since parties are commonly referenced by their acronym in Germany (e.g., “FDP” is the prevalent notion for the party “Freie Demokratische Partei”), we only consider those acronyms. In this paper, we write the acronym of a party  $P$  as  $acr(P)$ . Furthermore, we consider for each party a set of prominent party members  $M(P) = \{m_{P,1}, \dots, m_{P,l}\}$ . We use the notation  $m_{P,i} \sqsubseteq T(d)$  if the full name (first, middle and last name) of a party member is contained in the title of an article ( $m_{P,i} \sqsubseteq F(d)$ ,  $m_{P,i} \sqsubseteq K(d)$  for full texts and keywords respectively).

### 3.2 Coverage Bias Metrics

The first group of metrics is concerned with the coverage political parties receive. These measures should be interpreted in comparison to the values for other political parties. Therefore we apply standard normalization procedures. Given any raw measure  $\overline{M}_x(P, N)$ , the normalized measure  $\widehat{M}_x(P, N)$  is computed as  $\widehat{M}_x(P, N) = \frac{1}{Z} \cdot \overline{M}_x(P, N)$ , with the normalization constant  $Z = \sum_{P' \in \mathcal{P}} \overline{M}_x(P', N)$ . Additionally, differences between newspaper are easier to spot when only the deviation for this newspaper in comparison to the average values for the complete set of newspapers  $\mathbb{N}$  are displayed:  $M_x(P, N) = \widehat{M}_x(P, N) - \frac{\sum_{N \in \mathbb{N}} \widehat{M}_x(P, N)}{|\mathbb{N}|}$ . Please note that our measures do not consider multiple occurrences of acronyms or names and count every article only once.

#### 3.2.1 Party as Main Article Topic

A first group of measures indicates how often a party appears as the main topic of an article. In that direction,  $M_{Title}(P, N)$  describes how often political parties appear in the titles of a newspaper’s articles:

$$\overline{M}_{title}(P, N) = |\{d \in N : acr(P) \sqsubseteq T(d)\}|.$$

Since we are analyzing a corpus of online news, each article is associated by its publisher with certain keywords in the *HTML* header section. These keywords can also be used to identify articles, which are directly concerned with a political party. The respective measure  $M_{keywords}$  is computed as:

$$\overline{M}_{keywords}(P, N) = |\{d \in N : acr(P) \sqsubseteq K(d)\}|.$$

### 3.2.2 Party and Party Member Mentions

Further measures are concerned with the overall coverage political parties receive within the full text of the articles. In that direction, we count the number of distinct articles that contain a reference to a political party.

$$\overline{M}_{Full}(P, N) = |\{d \in N : acr(P) \sqsubseteq F(d)\}|.$$

As an extension, we do not consider references to the party itself, but to prominent party members, i.e., all parliament members of the respective party. Then, we count the proportion of articles in each newspaper that contain the exact name (first, middle and family name) of a prominent member. Using the exact name ensures that members are detected with high precision.

$$\overline{M}_{Full\_mem}(P, N) = |\{d \in N : (\exists m_{P,i} \in M(P) : m_{P,i} \sqsubseteq F(d))\}|.$$

## 3.3 Statement Bias Metrics

The second group of metrics is designed to indicate statement bias, i.e., (un-)favorable reporting about specific political positions.

### 3.3.1 Sentiment Analysis

We use state-of-the-art sentiment analysis for German language and identify for a neighborhood, i.e., four words before and four words after a mentioning of a party  $P$ , the mood  $S(N, P)$ . All party mentions are classified by *SentiStrength* [15] for a scaled sentiment strength using the full text  $F(d)$  of an article. We configure SentiStrength for the syntax of the German language (negation words can occur after the sentiment). An overall sentiment towards a political party in a newspaper can then be computed as the sum of the individual mood ratings.

### 3.3.2 Vocabulary Similarity

Although party manifestos show in general a strong overlap in the used vocabulary, each party lays a distinct emphasis on specific ideological terms such as *freedom*, *solidarity*, *environment*, etc., cf. [10]. Common usage of the same terms in an online newspaper and a political party can indicate a related ideology. To capture such similarities, we first determine a list of keywords that specifically point at certain political orientations. For each party, the number of keyword occurrences in recent party manifestos is counted and stored in a vector. Analogously, another vector then provides the number of keyword occurrences in each online newspaper. A formal measure for the vocabulary similarity between an online newspaper and a party is then given by the *cosine similarity* of the corresponding vectors.

## 4. DATASET

For our analysis, we utilized a large dataset consisting of articles published on four of the leading online news sites in Germany, that is, *faz.net*, *spiegel.de*, *taz.de*, and *zeit.de*. We restrict our analysis to the time between 27th October 2009 and 22th October 2013 covering the 17th legislative session of the German federal parliament. Six main political parties sent members to the parliament in this time frame: the conservative sister parties *CDU* and *CSU*, the liberal party *FDP*, the greens (*Grüne*), the social-democratic *SPD* and the left-wing party *Linke*. The government in this time span

**Table 1: Dataset overview: The number of articles ( $|N|$ ), their average length ( $l$ , measured in words), the share of articles that mention at least one party ( $s_P$ ) or party member ( $s_M$ ) or either party or member ( $s_C$ ) in the full text, and the percentage of articles that mention a party acronym in the title ( $s_T$ ) or in the keywords ( $s_K$ )**

$N$	$ N $	$l$	$s_P$	$s_M$	$s_C$	$s_T$	$s_K$
faz.net	40,209	513.5	30.5%	24.5%	35.1%	4.1%	20.1%
spiegel.de	51,560	472.1	31.7%	30.6%	37.5%	6.0%	10.0%
taz.de	23,610	477.5	35.9%	27.6%	40.0%	2.5%	2.9%
zeit.de	17,926	445.5	36.6%	35.7%	43.2%	6.8%	32.3%
Average	33,326	477.2	33.7 %	29.6 %	39.0 %	4.9 %	16.3 %

was formed by a conservative-liberal coalition of the parties *CDU*, *CSU* and *FDP*, led by chancellor Angela Merkel.

For the analyzed time frame we retrieved all articles that have been published in the respective *politics* and *economics* sections. We parsed the article pages and extracted the *date*, the article *title*, the *full text*, and the *keywords* added by the publisher as meta-information. For each part of the article, we identified mentions of the parties and party members that are also members of the parliament, see also Section 3.

Table 1 shows dataset characteristics including the share of articles that contain references to parties or party members as defined in section 3. Overall, our dataset features more than 130,000 newspaper articles containing more than 62,000,000 words. For each of the four online newspapers more than 370 articles are available on a monthly average. However, not all these articles can be easily associated with party politics as the dataset also includes other articles on e.g., economic topics or foreign politics. On average, only 39% of articles mention a party either directly or indirectly, i.e., by mentioning a prominent party member. From the analyzed news sites, *zeit.de* published the smallest amount of articles in the analyzed time frame, but it has the highest percentage of articles that mention a party.

*taz.de* started adding *keywords* only after February 2012. As a result only 2.9% of *taz.de* articles contain a party acronym in the *keyword* meta-information. Also care must be taken when judging results for keywords since we found that some articles are missing important *keywords*.

For vocabulary analysis, we relied on additional data from previous research in political science: Pappi, Seher and Kurella provide a short list of key vocabulary (such as *freedom*, *solidarity*, etc.) including the number of mentions in the respective election programs<sup>1</sup> aggregated from 1990 until 2009 [10].

## 5. RESULTS

This section presents experimental results of the described measures on the German online newspaper dataset.

### 5.1 Coverage Bias

In the following, we present results for the measures that indicate coverage bias, cf. Section 3.2. For these measures, the overall values for each party and online newspaper over the complete time interval are reported. In order to additionally test the statistical significance of our findings, we calculate

<sup>1</sup>Manifestos for the *CSU* have been excluded, since they have been identical to those of the *CDU* for most elections.

**Table 2: Differences in coverage: acronyms in titles.**(a) Articles with party acronyms in title ( $\overline{M}_{title}$ ).

$N$	CDU	CSU	FDP	Grüne	Linke	SPD
faz.net	315	167	481	213	57	543
spiegel.de	605	296	848	355	266	1,007
taz.de	104	28	120	127	99	146
zeit.de	224	118	381	147	90	369

(b) Normalization over every online newspaper ( $\widehat{M}_{title}$ )

$N$	CDU	CSU	FDP	Grüne	Linke	SPD
faz.net	17.7%	9.4%	27.1%	12.0%	3.2%	30.6%
spiegel.de	17.9%	8.8%	25.1%	10.5%	7.9%	29.8%
taz.de	16.7%	4.5%	19.2%	20.4%	15.9%	23.4%
zeit.de	16.9%	8.9%	28.7%	11.1%	6.8%	27.8%
Average	17.3%	7.9%	25.0%	13.5%	8.5%	27.9%

(c) Differences in coverage using titles ( $M_{title}$ ).

$N$	CDU	CSU	FDP	Grüne	Linke	SPD
faz.net	+0.4%	+1.5%	+2.1%	-1.5%	<b>-5.2%</b>	+2.7%
spiegel.de	+0.6%	+0.9%	+0.1%	<b>-3.0%</b>	-0.6%	+1.9%
taz.de	-0.6%	<b>-3.4%</b>	<b>-5.8%</b>	<b>+6.9%</b>	<b>+7.4%</b>	-4.5%
zeit.de	-0.4%	+1.0%	+3.6%	-2.4%	-1.7%	-0.1%

**Table 3: Differences in coverage: acronyms in keywords ( $M_{keywords}$ ).**

$N$	CDU	CSU	FDP	Grüne	Linke	SPD
faz.net	<b>+6.1%</b>	+1.1%	<b>+5.5%</b>	<b>-10.7%</b>	<b>-2.3%</b>	<b>+0.3%</b>
spiegel.de	-0.5%	+1.3%	-0.8%	-0.3%	+1.3%	-1.0%
taz.de	-7.5%	<b>-2.7%</b>	-6.2%	<b>+10.8%</b>	<b>+1.8%</b>	+3.8%
zeit.de	+1.9%	+0.3%	+1.6%	<b>+0.2%</b>	-0.8%	<b>-3.1%</b>

all metrics on a monthly basis in the considered time span. With the computed values, we conduct a *Student's t-test* for a single sample for each party and online newspaper and test if the mean value is significantly different from zero. We mark the *Bonferroni* corrected ( $m = 24$ ) significant values at a level of 0.05 bold in the result tables.

### 5.1.1 Articles about Parties

Table 2 shows results of analyzing party occurrences in article titles ( $M_{title}$ ). Using this measure, we also demonstrate the normalization procedures. First, the overall count of occurrences  $\overline{M}_{title}$  is displayed in Table 2(a). Then, this table is normalized to obtain the share of counts  $\widehat{M}_{title}$  for each party, see Table 2(b). Finally, the average value for each party over all online newspapers is subtracted from the individual newspaper values. Thus, the displayed values show how much more or less coverage compared to the average a party got in each newspaper. Due to constraint space, results for other measures are only reported after normalization.

Table 2(c) shows the differences of party mentions in article titles. The newspaper faz.net tends to slightly favor the governing parties CDU, CSU and FDP to the leftist Linke and Grüne. Results for taz.de indicate an opposite ideology. For the newspapers zeit.de and spiegel.de, the picture is less clear since there are few significant findings. This indicates that these newspapers are ideologically in the center of the analyzed spectrum of online newspapers. These findings are in line with the public perception of these newspapers.

Party mentions in HTML keywords, see Table 3, show overall similar results, but with unusual outliers, i.e., mentions of the party Grüne.

**Table 4: Differences in coverage: acronyms in full text ( $M_{Full}$ ).**

$N$	CDU	CSU	FDP	Grüne	Linke	SPD
faz.net	<b>+1.4%</b>	<b>+0.5%</b>	+0.3%	+0.0%	<b>-2.5%</b>	+0.3%
spiegel.de	-0.1%	<b>+1.0%</b>	+0.7%	<b>-1.0%</b>	-0.2%	-0.3%
taz.de	<b>-1.9%</b>	<b>-2.0%</b>	<b>-2.5%</b>	<b>+2.4%</b>	<b>+3.1%</b>	<b>+0.8%</b>
zeit.de	+0.7%	+0.5%	<b>+1.5%</b>	<b>-1.5%</b>	-0.4%	-0.8%

**Table 5: Differences in coverage: party members in full text ( $M_{Full.mem}$ ).**

$N$	CDU	CSU	FDP	Grüne	Linke	SPD
faz.net	<b>+3.7%</b>	+0.2%	-0.8%	<b>-1.4%</b>	<b>-1.6%</b>	-0.0%
spiegel.de	+1.8%	+0.4%	+1.0%	<b>-1.1%</b>	<b>-1.8%</b>	-0.4%
taz.de	<b>-6.5%</b>	-0.6%	<b>-2.3%</b>	<b>+4.2%</b>	<b>+4.3%</b>	+0.9%
zeit.de	+1.0%	+0.0%	<b>+2.1%</b>	<b>-1.6%</b>	-0.9%	-0.5%

### 5.1.2 Articles that Mention Parties

Table 4 is obtained by counting articles that contain the party acronym in the full text. Again, faz.net favors conservative parties over the Linke in terms of overall coverage. In contrast, taz.de favors the Linke and Grüne over the conservative and liberal parties. zeit.de and spiegel.de display slight bias towards the liberal FDP and against the Grüne.

Table 5 shows results for counting articles that contain a reference to a prominent party member in the full text (see Section 3.2.2). The obtained values support the observations for Table 4. However, the deviations on party member level seem to be even more significant in comparison to the deviations for party acronyms.

## 5.2 Statement Bias

Regarding statement bias, we first investigate results for sentiment analysis. Table 6 shows the average scaled sentiment strength (see Section 3.3.1) of all parties in each online newspaper. As we can observe from the table, the average party scores do not deviate significantly from zero. Based on these scores, there is a slight tendency for all newspapers to present the CDU in a slightly positive and the FDP in a slightly negative way. This could be inferred from the fact that the party got entangled in several affairs in the considered time interval. faz.net reports overall slightly more positive about parties, in particular the conservative ones CDU and CSU. However, in summary results for the sentiment analysis are inconclusive. In [16] Atteveldt et al. report similar findings for Dutch newspapers. That is possibly due to the fact that party bias is not expressed bluntly in newspapers, but in a more indirect way, indicating the need for more sophisticated specialized techniques for sentiment analysis.

Finally, results for the similarity of key vocabulary in articles and party manifestos are shown in Table 7. It can be seen that the usage of key terms for the faz.net is more similar to governing coalition parties CDU and FDP than it is for other newspapers, especially compared to spiegel.de and taz.de. By contrast, party manifestos of Grüne and Linke are more

**Table 6: Average sentiment scale for every online newspaper and party.**

$N$	CDU	CSU	FDP	Grüne	Linke	SPD
faz.net	0.002	0.067	-0.013	0.002	0.005	0.014
spiegel.de	-0.026	0.028	-0.057	0.015	-0.065	0.021
taz.de	-0.018	0.041	-0.034	-0.020	-0.014	0.002
zeit.de	-0.023	0.066	-0.061	-0.014	-0.048	0.011

**Table 7: Similarity of the distribution of key vocabulary between online newspaper articles and party manifestos (possible range 0 to 1).**

N	CDU	FDP	Grüne	Linke	SPD
faz.net	0.863	0.853	0.731	0.794	0.838
spiegel.de	0.808	0.802	0.756	0.831	0.833
taz.de	0.756	0.774	0.759	0.835	0.802
zeit.de	0.846	0.842	0.787	0.845	0.864

dissimilar. Also for other newspapers interesting results can be observed: for taz.de, large distances to CDU and FDP vocabularies are evident. Results for zeit.de show higher similarities to the left parties SPD, Grüne, and Linke than other newspapers. Vocabulary used in party manifestos of Grüne is most dissimilar to articles in the analyzed online newspapers. Overall, the results point in a similar direction about the political orientation of online newspapers as findings with coverage bias measures.

## 6. DISCUSSION

It is important to notice that the proposed measures for coverage bias detection analyze a set of online newspapers comparatively, that is, the obtained results should only be interpreted in relation to the other analyzed online newspapers. As such, our method is highly dependent on the selection of online newspapers, e.g., a moderate paper would appear to be leftist when compared only to conservative ones.

Counting mentions of party acronyms and member names is subject to some issues: parties in government tend to generate more news and are differently framed depending on the party status [6]. Additionally, acronyms and names might be ambiguous. However, these factors influence all online newspapers in the same way and consequently don't distort the comparative analysis significantly.

For the proposed measures, different variations could additionally be considered. For example, the measures count each party only once for each article. Instead, each occurrence could also be counted individually. Since these variations lead to very similar results in our initial experiments, we do not report them in this paper. Several metrics proposed in this paper expose similar tendencies, cf. Section 5. Thus, not every measure is crucial on its own. However, by considering a broad spectrum of different measures, a more stable picture of the overall political orientation emerges.

## 7. CONCLUSIONS

This paper was concerned with the large-scale analysis of online newspapers. In that direction, we discussed several automatically computable metrics that indicate potential bias towards a political party. These considered the mentions of parties and party members in the title, the keyword and the full text of articles, the sentiment in the direct neighborhood of article mentions, and the vocabulary used in online newspapers in comparison to party manifestos. Using these metrics, we investigated a large dataset containing all articles from the politics and economics sections from four of the leading German online news sites, that is, faz.net, spiegel.de, taz.de, and zeit.de over a four-year period. As a result, we were able to detect statistically significant reporting differences across the analyzed online newspapers.

In the future we want to extend our work in different directions. First, we will enlarge our dataset by adding more news sites and by considering longer time intervals. Furthermore, we aim at developing more sophisticated measures by integrating more data sources such as political speeches. Additionally, while this paper focused on bias towards political parties, similar techniques could be used to study other types of biases, for example, gender bias or bias against racial minorities. Finally, a comparative analysis of newspaper sites with social media sources such as Twitter is an interesting topic of future work.

## 8. REFERENCES

- [1] D. D'Alessio and M. Allen. Media bias in presidential elections: a meta-analysis. *Journal of communication*, 50(4):133–156, 2000.
- [2] M. Gentzkow and J. M. Shapiro. What drives media slant? Evidence from US daily newspapers. *Econometrica*, 78(1):35–71, 2010.
- [3] J. Grimmer and B. M. Stewart. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 2013.
- [4] T. Groseclose and J. Milyo. A measure of media bias. *The Quarterly Journal of Economics*, pages 1191–1237, 2005.
- [5] D. Hillard, S. Purpura, and J. Wilkerson. Computer-assisted topic classification for mixed-methods social science research. *Journal of Information Tech. & Politics*, 4(4):31–46, 2008.
- [6] G. Hirst, Y. Riabinin, J. Graham, and M. Boizot-Roche. Text to ideology or text to party status? *From Text to Political Positions: Text analysis across disciplines*, 55:93, 2014.
- [7] D. E. Ho, K. M. Quinn, et al. Measuring explicit political positions of media. *Quarterly Journal of Political Science*, 3(4):353–377, 2008.
- [8] M. Laver and K. Beniot. Estimating Irish party position using computer wordscoring: the 2002 elections. *Irish Political Studies*, 17(2):97–107, 2003.
- [9] W. Lowe, K. Benoitt, S. Mikhaylov, and M. Laver. Scaling policy preferences from coded political texts. *Legislative Studies Quarterly*, 36(1):123–155, 2011.
- [10] F. U. Pappi, N. M. Seher, and A.-S. Kurella. Das Politikangebot deutscher Parteien bei den Bundestagswahlen seit 1976 im dimensionsweisen Vergleich, 2011.
- [11] K. M. Quinn, B. L. Monroe, M. Colaresi, M. H. Crespin, and D. R. Radev. How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54(1):209–228, 2010.
- [12] J. B. Slapin and S.-O. Proksch. A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, 52(3):705–722, 2008.
- [13] G. H. Stempel. The prestige press in two presidential elections. *Journalism & Mass Communication Quarterly*, 42(1):15–21, 1965.
- [14] G. H. Stempel and J. W. Windhauser. Coverage by the prestige press of the 1988 presidential campaign. *Journalism & Mass Communication Quarterly*, 66(4):894–919, 1989.
- [15] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558, 2010.
- [16] W. Van Atteveldt, J. Kleinnijenhuis, N. Ruigrok, and S. Schlobach. Good news or bad news? conducting sentiment analysis on dutch text to distinguish between positive and negative relations. *Journal of Information Technology & Politics*, 5(1):73–94, 2008.
- [17] B. Yu, S. Kaufmann, and D. Diermeier. Classifying party affiliation from political speech. *Journal of Information Technology & Politics*, 5(1):33–48, 2008.