Bias-aware News Analysis using Matrix-based News Aggregation[1]

Felix Hamborg, Norman Meuschke, Bela Gipp

University of Konstanz, Germany
`firstname.lastname@uni-konstanz.de`

Abstract. Media bias describes differences in the content or presentation of news. It is a ubiquitous phenomenon in news coverage that can have severely negative effects on individuals and society. Identifying media bias is a challenging problem, for which current information systems offer little support. News aggregators are the most important class of systems to support users in coping with the large amount of news that is published nowadays. These systems focus on identifying and presenting important, common information in news articles, but do not reveal different perspectives on the same topic. Due to this analysis approach, current news aggregators cannot effectively reveal media bias. To address this problem, we present matrix-based news aggregation, a novel approach for news exploration that helps users gain a broad and diverse news understanding by presenting various perspectives on the same news topic. Additionally, we present NewsBird, an open-source news aggregator that implements matrix-based news aggregation for international news topics. The results of a user study showed that NewsBird more effectively broadens the user's news understanding than the list-based visualization approach employed by established news aggregators, while achieving comparable effectiveness and efficiency for the two main use cases of news consumption: getting an overview of and finding details on current news topics.

Keywords: Media Bias, News Aggregation, Frame Analysis, Google News.

1 Introduction

The coverage of media outlets often exhibits *media bias*, e.g., due to political interference, lobbyism, or ideological focus [48]. Not only developing or autocratic countries, but also industrialized, democratic nations are subject to media bias. For instance, in the U.S., six corporations control 90% of the media [11], which results in a high chance of media manipulation [15, 58]. Trust in media is at a historical low, e.g., less than half of U.S. readers trust media and think it is objective [19].

Table 1 shows the headlines of two related news articles from November 7, 2014, during the Ukraine crisis. While Western media, such as CNBC, reported that Russian

---

[1] Part of the research described in this article has been published in the proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries 2017 [26].

tanks crossed the Ukrainian border, Russian media, such as RT, primarily portrayed these reports as false claims or did not mention the event. The content and tone of the articles differed just as strongly as the headlines suggest. One can assume that readers' perception of the event will differ significantly depending on which article they read.

Table 1. Different headlines for the same event.

Source	Headline
CNBC [62]	Tank column crosses from Russia into Ukraine: Kiev military
RT [61]	Moscow to Kiev: Stick to Minsk ceasefire, stop making false invasion claims

Media bias has many, severe effects, which Berhardt et al. discuss in detail [6]. A 2003 survey [37] showed a particularly troubling effect of media bias. The survey found significant differences in the presentation of information on the Iraq war by US television channels. The identified media bias apparently affected the news understanding of the channels' viewership. Fox News viewers were most misinformed about the Iraq war. Over 40% thought that weapons of mass destruction had been found in Iraq – a false claim the US government used as a justification for the war.

Although a rapidly increasing amount information from around the world is available online and often at no cost, many news readers only consult a small subset of news sources [47]. Reasons include the overwhelming number of sources, language barriers, or simply habit. These and further factors can cause a *narrow news perspective* [45], and thus a skewed or incomplete perception of information.

News aggregators, such as Google News, are information systems that focus on enabling news consumers to quickly and conveniently overview the news landscape, and explore important topics. As we show in Section 2, related systems in the field that we term *bias-aware news analysis* specifically aim to reduce media bias by finding different perspectives on the same topic [46, 51, 52]. However, due to limitations of underlying *natural language processing* (NLP) methods, these systems suffer from practical limitations, e.g., relying heavily on user feedback.

This article presents *matrix-based news aggregation* (MNA) as an analysis and visualization approach that enables users to explore both common and different information in related articles. Our main goal is to reduce the effects of media bias by broadening the users' news understanding, i.e., presenting different, possibly biased perspectives on a topic. Additionally, we present our news aggregator *NewsBird*, which exemplifies MNA for international news topics.

NewsBird addresses the information need of users, who want to quickly get an overview of commonalities and differences in news coverage on international events and topics within a custom date range. For this purpose, NewsBird structures news articles in a two-dimensional matrix, whose elements show the primary news perspective of one country (row) about another country (column). We refer to the primary news perspective as what readers of one country would typically read in most of the country's

news outlets, i.e. the mainstream media, on a specific topic or country. Users can also specify a custom query to view only news coverage on specific topics.

We structure the remainder of this article as follows. Section 2 gives an overview of the research on media bias and news analysis, particularly on news aggregation. Section 3 introduces MNA and describes its analysis approach. MNA lays the groundwork for the news aggregator NewsBird, which we present in Section 4. Section 5 describes the findings of our evaluation. Section 6 summarizes the capabilities of MNA and NewsBird, and presents our plans for future research.

2   Background and Related Work

2.1   Differences in News

Media bias can significantly change people's awareness and perception of topics. This change can become critical for issues with high social impact, such as elections [6] or people's attitude towards war (cf. Section 1). Reasons for biased news coverage include internal factors, such as that news consumers mostly want to receive confirmatory information [21, 44], and external factors, such as that journalists emphasize privately obtained information [5], or that governments influence publishers in their favor [7]. Aside from such intentional influences, unconscious influences called *news values* also affect the news production process. For instance, an accident with ten fatalities will be more important to readers living closer to the location of the event, hence raising the likelihood of the event being reported by local news outlets [28].
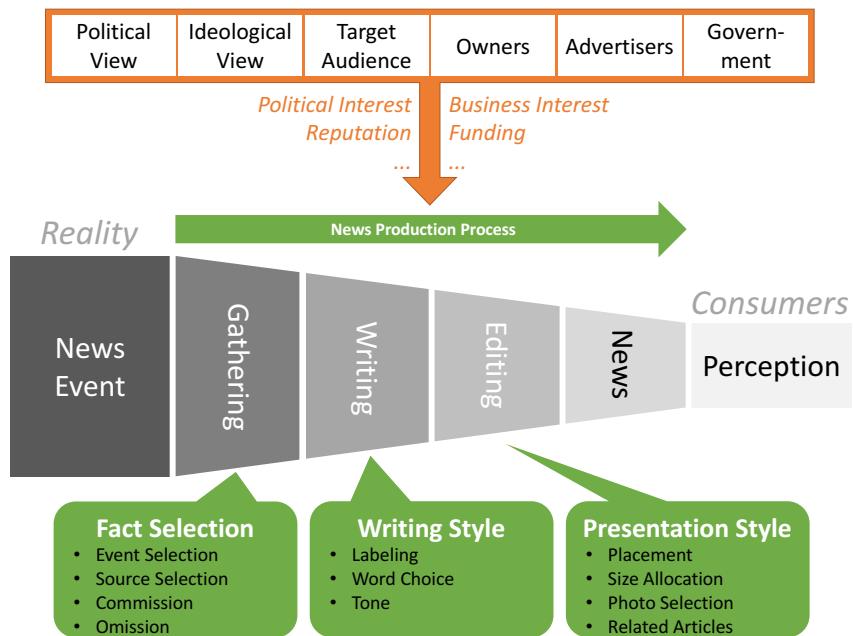

Fig. 1. Reasons and forms of media bias. Based on: [51].

Fig. 1 depicts different forms of media bias that can occur in the news production process, in which publishers transform an actual event into a news story [51]. During the initial *gathering* step of the process, journalists select events, sources, and the facts they want to present. This selection biases the resulting news story. During the *writing* step, journalists can affect the reader's perception of a topic, e.g., through word choice. The author can choose a positively or negatively connotated word to refer to an entity, such as "coalition forces" vs. "invasion forces", or vary the credibility ascribed to the source [3, 21, 49]. The third step, *editing*, determines the (visual) presentation of an article, e.g., through placement. For instance, a front-page article receives the most attention. Finally, consumers read the news. Reading may also yield different perceptions of the event [4, 59], but this influence is beyond the focus of this article.

In conclusion, media bias is a structural, often intentional, flaw inherent to news publishing [14] and can critically impact people's opinions and decisions. Before Section 2.3 describes related work on bias-aware news analysis to reduce the effects of media bias, Section 2.2 introduces some background knowledge on news aggregators and their underlying analysis methods. Section 2.4 then describes the technical challenges that lead to poor media bias identification and bias reduction in such systems. We discuss the findings of our literature review in Section 2.5.

2.2 News Aggregation

News aggregation is a state-of-the-art approach to let readers overview the large amount of news that is produced nowadays. The analysis workflows of most news aggregators find the most important news articles and summarize them for users. This typically involves the following steps [16]:

1. **Data gathering**, i.e., crawl articles from news websites.
2. **Article extraction** from raw website data.
3. **Grouping**, i.e., find and group related articles about the same topic or event.
4. **Summarization** of related articles.
5. **Visualization**, e.g., present the most important topics to users.

For the first two steps, *gathering* and *extraction*, established and reliable methods are available, e.g., as part of web crawling frameworks [43]. Articles can be extracted using naive approaches, such as website-specific wrappers [50], or more generic methods using content heuristics [34].

Integrated aggregation systems combine the first two steps and provide additional functionality. For example, *news-please* supports full website extraction, i.e., collecting all articles of a news outlet by providing only a root URL [27].

The main objective of the *grouping* step is to identify topics and use them to categorize articles. To accomplish these tasks, established systems typically employ topic modeling, e.g., using *Latent Dirichlet Allocation* (LDA) [9] as for instance applied in the Europe Media Monitor [8], or clustering methods, such as hierarchical agglomerative clustering (HAC) as used in NewsCube [51] and Newsblaster [40]. Articles are then summarized using a broad spectrum of methods ranging from simple TF-IDF-based scores to complex approaches considering redundancy and order of appearance, such as *MEAD* [54].

Established news aggregators, such as *Google News*, have similar user interfaces, which typically show a list of topics ordered by relevance to the user query or by topic frequency. For each topic, such news aggregators select the most representative article during the *summarization* step, and *visualize* the results for the user by displaying an article's headline and lead paragraph, as well as related articles.

Some systems use less conventional user interfaces. For instance, *newsmap* features a two-level treemap to show news categories and topics [60]. *Hiérarchie* shows a topic hierarchy in a sunburst diagram to let users explore different semantics of a topic [57]. Aside from commercial news aggregators, the scientific community has developed various approaches that analyze and aggregate news. Most relevant to our goal of broadening a user's news understanding are *Newsblaster*, which is one of the first academic news aggregators [40], the *Europe Media Monitor*, which improves automatically aggregated news through manual revision [8], and *PNS*, a news aggregator that provides user personalization [50].

The presented analysis workflow enables news aggregators and other news analysis systems [8, 40, 50] to process the vast amount of news produced every day. Their large user base as well as the retrieval performance and the usability scores such systems achieved in scientific evaluations [8, 50] indicate the maturity of the systems and the analysis approach.

However, no news aggregator focuses on revealing differences between related articles [51] and few systems offer functionality that could be used for this purpose (see Section 2.4). Thus, users of established news aggregators are subject to media bias [10].

2.3   Bias-aware News Analysis

Traditional efforts to broaden readers' understanding of news rely on manual analysis and presentation. Popular presentation formats include the opposite editorial, in which two or more authors argue in favor of opposing positions on a topic, and the press review, in which news outlets present a summary of statements of different publishers on the same topic.

Systems to support the task that we name *bias-aware news analysis* aim at finding and presenting different perspectives on a topic. *NewsCube* uses so called *aspect-level browsing* to enable users to view different perspectives on political topics [51]. An aspect represents a semantic component of a news topic. The approach follows the workflow described in Section 2.2, but includes a novel grouping step: NewsCube extracts aspects from each article using keywords and syntactical rules. The system then weighs aspects according to their position in the article using the *inverted pyramid* concept: the earlier an aspect appears in the article, the more important the system considers the aspect. NewsCube then performs HAC to group related articles. The offered visualization is similar to the visualizations of other aggregators, but additionally shows different aspects of a selected topic. The experiments of Park et al. showed that NewsCube users became aware of such perspectives, and subsequently read more articles containing the respective aspects [51].

NewsCube 2.0 uses a manually curated list of publishers to show the perspectives on a selected topic. The system also enables users to collaboratively extend and improve

the assumed publisher perspectives [52]. The evaluation of NewsCube 2.0 showed that the diversity and usefulness of information highly depend on the quality of users' feedback and can vary strongly if only partially related articles are presented as related.

Sideline uses blogs that were manually classified according to their political orientation to identify different perspectives on political topics [46]. To assess an article's orientation, Sideline determines how many blogs of each orientation link to the article. The approach measurably reduces the readers' tendency of sharing the perspective that is most frequently presented. A related approach proposed by Park et al. uses the sentiment of readers' comments to estimate the political slant of a news article [53].

Comparative or *contrastive summarization* methods aim to summarize both common and unique sections in a set of documents. Newsblaster, one of the first news aggregators, supports a basic comparative summarization that shows for two groups of news articles the top-ranked summary sentences according to a summarization score [16]. While the summaries enable the user to get an overview about both article groups, comparison is rather difficult as there is no alignment of comparable phrases or summaries. A more recent approach aligns comparable phrases, but processes topics instead of articles reporting on the same topic [32].

2.4   NLP Methods in News Analysis

A major reason for the inability of today's news aggregators to identify media bias is the poor performance of current NLP methods in identifying semantic differences in news [51]. Classic NLP techniques typically rely on statistics and are "[...] just a first step towards natural language understanding" [12]. For instance, even clearly opposing articles, such as the two articles in Table 1, have a high cosine similarity when expressed as TF-IDF vectors, because articles on the same topic typically share many topic-specific keywords. Another example of wording with opposing semantics, but high TF-IDF similarity are references to the Iraq war. While Western media typically referred to the conflict as "War in Iraq" or "Iraq war", Iraqi media used "War on Iraq".

The semantic analysis of news using current NLP methods is particularly challenging, because semantic differences in news are often encoded subtly due to the requirement for journalistic objectivity [20]. While sentiment analysis yields good results for texts in which authors explicitly state their opinion, such as product reviews [31], the results for news texts are not satisfactory [49]. Employing sentiment analysis to find articles that differ in their coverage, e.g., articles that report either positively or negatively on a politician, typically yields poor results.

While some approaches that aim at revealing differences among news or more generally text documents exist, these approaches suffer from the limitations of NLP methods. For example, comparative summarization methods list the most important common and differing information in multiple news articles [16] or topics [32]. However, the quality of section alignment requires further improvement before these techniques allow for an effective comparison [16]. Another approach uses recursive topic modeling to find different (semantic) components of a topic [57]. The resulting topic components are not always meaningful, but often represent artificial subtopics. Recently, researchers proposed language models to identify biased language in particular types of

text, such as Wikipedia articles [55]. To our knowledge, however, there are no such models for news articles, likely due to the subtlety of biased language in news coverage.

In summary, the vast majority of methods for the automated analysis of text semantics are highly domain-specific or use case-specific, require much manual effort, e.g., for creating suitable ontologies, or perform poorly in finding meaningful differences in news articles. Thus, we conclude that the exclusive utilization of state-of-the-art NLP techniques is not sufficient to identify the subtle semantic differences in news.

2.5 Summary of Related Work

Our review of related work showed that media bias can cause a strong misperception of information and events, especially when the presentation of information is intentionally biased. Readers can reduce the effects of media bias by reading articles that present different perspectives on an event. Yet, most people consult very few news sources. Established news aggregators present information that related news articles have in common, instead of revealing information that differs between the articles. There are approaches that can reduce the effects of media bias by broadening users' understanding of news topics. However, these approaches suffer from practical limitations, such as being restricted to the analysis of one news category [46, 51], requiring manually created knowledge bases [52], and being fine-tuned for specific analysis tasks.

This article presents an extended version of our preliminary paper on MNA and NewsBird [26]. In addition to our previous work, we describe a user study, which we conducted to quantitatively and qualitatively evaluate NewsBird's effectiveness and efficiency to reveal differences in news coverage (Section 5). Furthermore, we conceptually compare advantages and disadvantages of NewsBird with established news aggregators (Section 4.7). We also describe important building blocks of MNA in more detail than was possible in the conference publication. For example, in Section 4.1 we discuss the bias induced by the skewed language distribution in our dataset prior to the machine translation of all articles to English. We also discuss alternative visualization concepts suitable to show differences in news coverage (Section 4.6). Lastly, we extend the discussion of advantages and disadvantages of MNA, NewsBird, and the employed techniques (Section 6).

3 Matrix-based News Aggregation

Matrix-based news aggregation (MNA) is a generic news exploration approach that follows the analysis workflow of news aggregators explained in Section 2.2, but includes an additional step before the grouping step.

MNA reveals different perspectives in news by structuring news articles in a two-dimensional matrix, whose elements show what entity i (row) states about entity j (column). The dimensions of the matrix can encode arbitrary entities ranging from politicians to the media landscape of countries or regions. Rows and columns can encode different entity types as depicted in Fig. 2. For instance, the matrix elements could show

the main content that the media in one country (row) report about another country (column), i.e., what a reader from one country would typically read about another country.
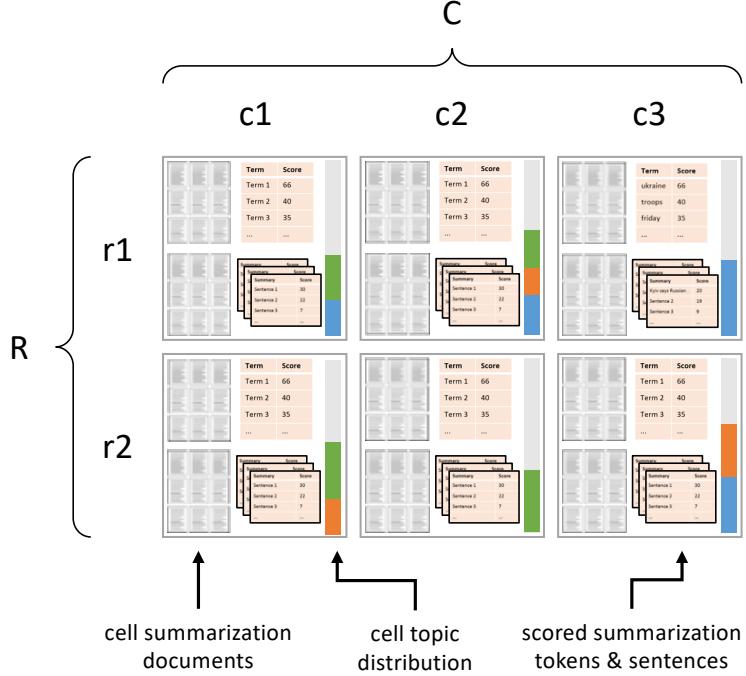

Fig. 2. Organization of articles, topics, and summaries.

To reveal the relations between the chosen entities, MNA groups news articles into the cells of a matrix created upon user request. We call articles that have been assigned to a cell *cell documents*. MNA then summarizes the topics of the articles in each cell. For example, for an international news topic, such as an armed conflict, spanning a matrix over countries (publisher country × mentioned country) will likely yield highly diverse content in the resulting cells, particularly if the countries involved in the conflict are included. The example of the Ukraine crisis presented in Section 1 demonstrates the idea of the approach. We showed that countries have very different perspectives on the same event (cf. Table 1).

Fig. 3 depicts the analysis workflow of MNA. The first two steps of the approach – *data gathering and article extraction* – create or update the database of news articles either as a one-time or as a recurring process. This article focuses on describing of the novel components of MNA. For the following description, we thus assume that a dataset of news articles exists.

To start the analysis, the user must define the *analysis scope* (see also Section 4.2), primarily by specifying the query date and the dimensions. For getting an overview of today's events and getting more detailed information for a specific event, which are the most common use cases in news consumption, the user is not required to enter information, since MNA provides default values for this purpose. For instance, a reader from

a European country is assumed to be primarily interested in events and media coverage in Western countries occurring on the current day.
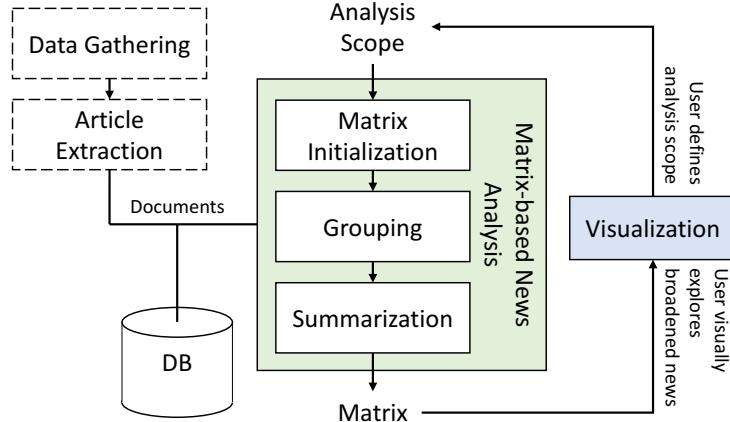

Fig. 3. MNA analysis workflow.

The first step in the analysis workflow, *matrix initialization*, spans a matrix over the two chosen dimensions, and finds the cell documents for each cell. For the example of news coverage on the Ukraine crisis, the cell of the publisher country Russia (row) and the mentioned country Ukraine (column), hereafter denoted with RU-UA, contains all articles that have been published in Russia and mention Ukraine.

The *grouping* step collects related articles, i.e., articles that report on the same topic. MNA uses the documents in all cells of the matrix to find topics.

Finally, the *summarization* step generates the following three summaries:

1. *topic summary* for each topic: MNA considers all documents containing the topic to create this summary.
2. *cell summary* for each cell: MNA considers all documents in a cell to create this summary.
3. *cell topic summary* for each topic present in one cell: MNA considers all cell documents containing the topic to create this summary.

MNA yields the matrix depicted in Fig. 2 – each cell contains one or more weighted topics and the corresponding summaries.

Users can control the analysis workflow, especially by defining the dimensions of the matrix. MNA provides a set of default dimensions, but also lets users add and interactively refine dimensions. This feature offers two main advantages over existing approaches: while MNA does not require user input to support revealing different perspectives, users can improve the analysis results by incorporating their knowledge. For instance, a user might be aware of differences in the coverage of certain news outlets, and hence could span a matrix over these outlets and the countries they mention.

Compared to established approaches, the workflow of MNA enables a flexible analysis of different news categories, various analysis questions, and lets users control the analysis workflow by incorporating their domain-knowledge.

4 System Description: NewsBird

NewsBird is an open-source news aggregator that enables bias-aware news analysis. Currently, NewsBird focuses on international news. To overcome the issues of media bias described in Section 2.1, NewsBird implements MNA as shown in Fig. 3. Our description of NewsBird follows the MNA analysis approach consisting of data gathering and article extraction, matrix initialization, grouping, summarization, and visualization (cf. Section 3).

NewsBird is openly available under an Apache 2 license at:
https://github.com/fhamborg/NewsBirdServer

4.1 Data Gathering and Article Extraction

The current version of NewsBird uses fixed datasets instead of performing data gathering and article extraction, since the system currently focuses on demonstrating the novel components of the MNA approach. The dataset that we used for devising MNA and implementing NewsBird originates from the Europe Media Monitor (EMM) [2] and consists of 1.6 million articles in more than 30 languages gathered from approx. 4,000 publishers from over 100 countries between October to November 2014.

Currently, NewsBird can only process documents in English. We made this design decision to cope with the limitations of NLP technologies we employ, such as topic modeling, which is significantly less reliable when performed across languages.

Despite the system's limitation to processing English texts, we did not simply exclude non-English documents to not bias the investigation. If we did, NewsBird could have no longer revealed the news perspective of "typical" readers in countries whose first language is not English. Most people read news articles in their first language. To illustrate this fact, Fig. 4 shows the share of English news articles published in different countries. The left side of Fig. 4 shows the ten countries that publish the largest number of English news articles. The right side of Fig. 4 shows the ten countries that exhibit the largest difference in the number of news articles published in English and the number of news articles published in the countries' first languages.
Countries on the right side of Fig. 4, such as France or Germany, publish a large share of all news articles, but most of them in languages other than English.

Systems that focus on bias-aware news analysis need to also analyze non-English articles to avoid missing a large portion of relevant information. Therefore, our dataset includes English and non-English articles from European countries. We translated all non-English articles in our dataset to English using a machine translation service[2].

The quality of machine-translated text is lower than the quality of manual translations, yet often high enough for IR tasks [16, 17] and sufficient for our purpose.

Each article in the dataset contains a title, a lead paragraph, content, i.e., the main text, publishing date, and other metadata. Since the dataset covers many sources from different countries, we consider it suitable for finding various, potentially contrary news perspectives for a given topic. We parsed the dataset and stored the resulting documents

---

[2] https://www.microsoft.com/en-us/translator/translatorapi.aspx

in an Apache Lucene index. Lucene performs state-of-the-art text preprocessing, such as tokenization, lowercasing, stop word removal, and stemming [30].
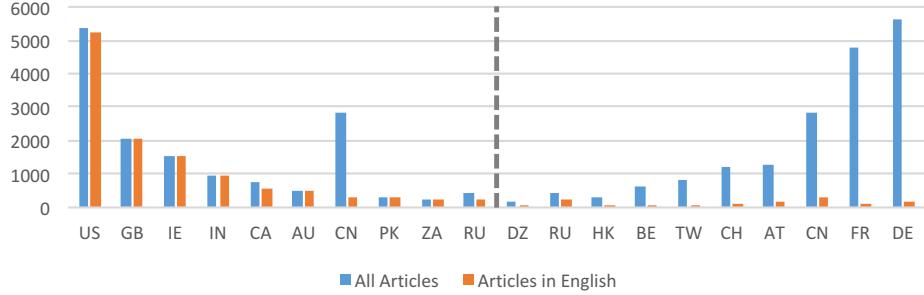

Fig. 4. Average number of all and English news articles published per day in different countries.

4.2 Analysis Scope

The analysis scope consists of a base query and an optional custom query. The *base query* specifies the date range to be analyzed and the two dimensions of the matrix. Each dimension consists of a list of values, e.g., the publishing countries. Specifying a date range enables users to also analyze news events in the past. Some established systems also offer this feature, but typically provide fewer analysis options for past news. For instance, Google News simply lists corresponding articles for a query in the past, rather than showing the main user interface that is exclusively available for events on the current day. The *custom query* allows users to enter keywords to restrict the analysis to a certain topic.

4.3 Matrix Initialization

NewsBird constructs the matrix for a given analysis scope, i.e., retrieves the specific values of each dimension, converts each value into a query constraint, and fills the cells in the matrix. For each cell, NewsBird constructs a query that is a conjunction of the base query and the *cell query*. The cell query will exclusively retrieve documents that meet the criterion specified by the dimension values. E.g., the cell query for the RU– UA use case will only retrieve documents published in Russia that mention Ukraine.

To form a representative sample of all cell documents, while limiting the time required for the subsequent topic extraction and summarization steps, the system retrieves $k_d = 100$ cell documents for each cell. Choosing a constant number of cell documents reduces the effects of selection bias on the topic modelling step, because each country equally contributes to the resulting topics, regardless of the number of articles that are published in a country. As shown in Fig. 4, the number of published news articles differs greatly among countries. For example, in our dataset, on average more than 1,000 articles per day are published in the U.S., whereas on average only 180 articles per day are published in Algeria. We found that choosing $k_d < 50$ significantly increases the

probability of missing important topics for countries that publish many articles. However, choosing large $k_d > 1{,}000$ proportionally increases the processing time needed for topic modeling and summarization. We achieved good results for topic modelling and summarization using $k_d = 100$. Choosing 100 cell documents is a large enough set to cover the most important topics, but small enough to achieve a good runtime performance. If no documents match a cell's query, the cell will be omitted from further processing and displayed without content in the visualization.

NewsBird currently supports the following dimension types:

1. **Publisher country**: documents published in a specific country. To determine the publisher country, we use the metadata available in the dataset.
2. **Mentioned country**: documents that mention a specific country. NewsBird checks if the specific country name occurs in the document's lead paragraph. Alternatives we explored include checking the title, which yielded low recall, and checking the main text, which yielded low precision. To increase the recall, we applied query expansion (using DBpedia and WordNet), but received mixed results, which is why query expansion is currently disabled.
3. **Time-range**: to explore how news coverage changes over the time. Users can select a granularity, e.g., daily, as well as a start date and end date for the analysis, e.g., Sept. 1 to Sept. 5, 2017. In this case, the resulting matrix allows users to analyze the development of topics between Sept. 1 to Sept. 5, 2017 by representing one day of news coverage in each row / column.

The two country dimensions enable users to comparatively explore international news topics. To support more news categories, additional dimensions can be easily added to the system (see Section 6).

4.4   Topic Extraction to Find Related Articles

NewsBird uses LDA to extract a list of topics from the matrix and assigns documents with identical topics to the same group. Particularly, NewsBird performs four subtasks. The first subtask is *text extraction*. We consider the full content (see Section 4.1) of each document, since LDA performs better for longer texts, as they have an increased chance of topic-defining term co-occurrences.

The second subtask is the actual *topic modeling*. The input to this subtask are the texts that have been extracted from all articles in all cells. The parameter configuration of LDA is crucial for the quality of topics. For a $m \times n$ matrix, we set the number of topics to mnk_t, where k_t controls the granularity of resulting topics. In all our tests, we achieved the best results with $k_t = 2$. This setting allows each cell to have at least one cell-specific topic, i.e., a topic that semantically represents the main content of the cell's documents. NewsBird performs 1,500 LDA iterations, the smallest number that yields stable results in our system. We set the Dirichlet hyperparameters $\alpha = \beta = 0.0001$ to stimulate the creation of cell-specific topics in each document and thus in each cell [22]. LDA then generates a list of topics and their weighted mappings to the cell documents. Finally, we average the weights of each cell's documents to obtain a weighted mapping of topics to cells.

The third subtask is *post-processing* of the resulting topics and mappings. Theoretically, each cell contains a weighted proportion of all topics, of which most have a weight of ≈ 0. Thus, we remove all improbable topics (weight < 0.2) from a cell. Similarly, we remove all improbable terms from a topic (weight < 0.002). Aside from cleaning improbable topics, we investigated methods to merge semantically overlapping topics. Since LDA is a generic topic modeling approach, not tailored to the characteristics of news articles, even the best performing LDA configurations in our tests yielded semantically overlapping topics. An example of two semantically overlapping topics is shown in Table 2. To our knowledge, no work has been published that investigates automatic topic matching for news articles. The only approach to topic matching we are aware of was proposed by Alexander and Gleiche, who use an interactive visualization to let users investigate and match topics. They employ a variant of the Cuthill-McKee algorithm [13] to group similar topics within a topic similarity matrix [1].

Table 2. Example of two semantically overlapping topics and their top-ranked LDA.

Topic	Top-ranked LDA Terms
Topic 1	ukraine, russia, troops, eastern, friday
Topic 2	ukraine, military, russia, border, tanks

We tested different methods to merge similar topics or discard all but one topic from a group of similar topics. A naive method is to measure the term overlap between two topics, e.g., two of five terms in Table 2 match. We found that this method has a good recall, but decreases precision strongly, i.e., too many non-similar topics are merged as well. If the merging threshold is more restrictive, recall decreases too much, because synonyms are not resolved. Due to the mixed results, NewsBird does not employ topic matching by default, but allows the user to enable topic matching and customize the matching parameters via the user interface. We discuss possible directions for future research to improve the topic matching procedures in Section 6.

The fourth subtask is *grouping* documents using the extracted topics. We do not use the original cell documents retrieved in the matrix initialization task, since they only satisfy the cell query. Instead, for each cell, we construct a *cell topics query* to find documents that satisfy the cell query and represent the previously extracted main topics of a cell. Expanding the search space is necessary, since the number of original cell documents containing the main topics is often low. The cell topics query consists of one weighted *topic sub-query* for each of the cell's topics. Each of the topic sub-queries consists of the $k_{\mathrm{t,top}}$ terms with the highest weight, of which at least $\frac{k_{\mathrm{t,top}}}{2}$ must be contained in a document to be retrieved. The resulting documents are called the *cell summarization documents*. We found setting $k_{\mathrm{t,top}} = 15$ to be a good trade-off between precision and recall. Users can modify all parameters in the visualization.

4.5 Summarization

NewsBird uses the sum of the TF-IDF and topic weights to compute the summary score for each sentence occurring in any of a cell's summarization documents. Our summary score is conceptually similar to MEAD [54], since we also use the TF-IDF scores of tokens to measure descriptiveness of a sentence and consider the topic distribution of a cell. For each cell, NewsBird computes separate summary scores for the title, the lead paragraph, and the main content of any of the cell's summarization documents (see Fig. 2) as follows. First, the system calculates the score of each *token w* occurring in the cell summarization documents D:

$$s(w, D) = \sum_{\{d \in D\}} TF - IDF(w, d, D) \tag{1}$$

where d is a single document in the set of documents D. After computing the scores of each token, NewsBird calculates the score of a *phrase* or sentence p as:

$$S(p, D) = \sum_{\{w \in p\}} s(w, D) \tag{2}$$

While the sentence score $S(p, D)$ represents the descriptiveness of a sentence p in its document set D [54, 56], we also compute the summary score $S(p, D, T)$ that additionally takes into account the sentences' relation to the cell's topics as:

$$S(p, D, T) = S(p, D)\alpha \sum_{\{t \in T\}} c(t) \sum_{\{w \in p\}} u(w, t) \tag{3}$$

where $c(t)$ is a function that returns the weight of a topic t in the current cell and $u(w, t)$ returns the weight of token w in topic t. We set the constant $\alpha = 9400$ to balance the non-normalized topic weights and TF-IDF scores.

To improve the quality of summaries, we adjust scores using rules suggested by Lin et. al [38], such as lowering the score if a sentence starts with a conjunction. Such sentences often refer to the previous sentence, without whom they cannot be interpreted by the user, e.g., "However, the politician disagreed with that statement." Also, we discard all sentences shorter than 30 characters to remove frequent, yet document-specific sentences, such as "Copyright by RT". By default, we only consider the first ten sentences of each document, since these convey the most descriptive information [51]. The limitation also increases the speed of computation. We then rank tokens and sentences of a cell according to their score.

In addition to the described cell summaries, NewsBird also summarizes each of the topics itself. Therefore, it produces one query for each topic.

4.6 Visualization

Different approaches are suitable to visualize the results of the MNA workflow. Hereafter, we discuss and explain the design choices we made for the visualization implemented in NewsBird. We distinguish possible visualization approaches into (1) *list-based*, (2) *time-based*, (3) *geo-based*, and (4) *relation-based*.

List-based visualizations are characteristic for the user interfaces of established news aggregators. Google News and similar systems show a list of topics and the most representative news articles for each topic. Commonly, news aggregators show details for the first article in the list, e.g., its headline, lead paragraph, and main image. This allows users to understand the topic that underlies the articles in the list. The list-based interfaces of popular news aggregators are intuitive and familiar to users. If additional information to enable bias-aware news analysis are integrated into the list-based visualization, we see this type of interface as a suitable and intuitive solution for our use case.

Fig. 5 shows our vision of an extended list-based topic visualization to support bias-aware news analysis. The upper left part of the view shows the same information that established news aggregators provide to users, i.e., the headline of the most representative article for the topic, an excerpt of its lead paragraph, and the article's main image. Additionally, the extended topic view shows a diversity-score indicating how controversial a topic is, i.e., how much differing information can be found within a group of related articles or between its sub-groups. A map shows the geographic distribution of the topic and uses glyph overlays to highlight regions in which the topic is particularly popular. In the lower part of the view, summaries of articles that represent the most divergent perspectives on the topic enable the user to quickly get a broader understanding of the topic. Unfortunately, current NLP techniques are not able to reliably select the most relevant and differing sub-groups within a set of topically related news articles (cf. Section 2.4), which is crucial to realize this kind of visualization.
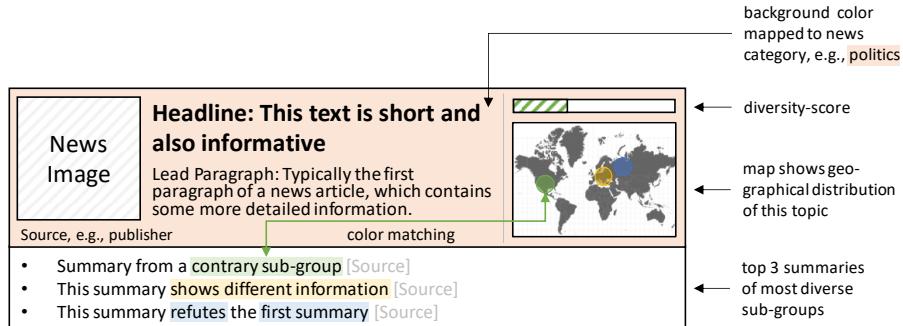

Fig. 5. News topic view for bias-aware news analysis. Source: [25].

Time-based news visualizations (cf. [35, 36]) visualize the temporal development of news topics, which is a promising option for bias-aware news analysis. For instance, a temporal visualization could show whether authors copy-edit from news agencies to support users in investigating the impact of specific news sources. However, this kind of analysis currently exceeds the capabilities of MNA, because, we currently do not compute article-to-article similarities.

Geo-based visualizations focus on displaying the geographic origin of data items. Visualizing this information is relevant to our current analysis approach, which considers countries to reveal different perspectives in news coverage. For international news coverage, such different perspectives often become apparent by organizing articles into

cells of a matrix spanned by publisher countries and mentioned countries (see Section 4.3). For other news categories, such as finance or domestic politics, a geo-based analysis and visualization are likely suboptimal. Therefore, we seek to develop a more generic visualization suitable for geographic and other groupings of news articles.

Relation-based visualizations, such as treemaps, sunburst diagrams, or matrices, are typically employed to display information that exhibit a hierarchical or graph-based structure. Matrix visualizations are more generic than sunburst diagrams or treemaps, which typically require some initial training. For our use case, matrix-based visualizations offer the additional advantage of a close link between the organization of news articles, topics, and summaries during the analysis and the visualization of this information for the user. Therefore, we decided to devise a matrix-based visualization for our prototype system, which we describe in the following.

NewsBird's matrix visualization enables users to explore the relations between two dimensions, e.g., what the media in one country state about another country. The first step in the exploration process is to define the analysis scope. After the system has completed the analysis, results are visualized as shown in Fig. 6.
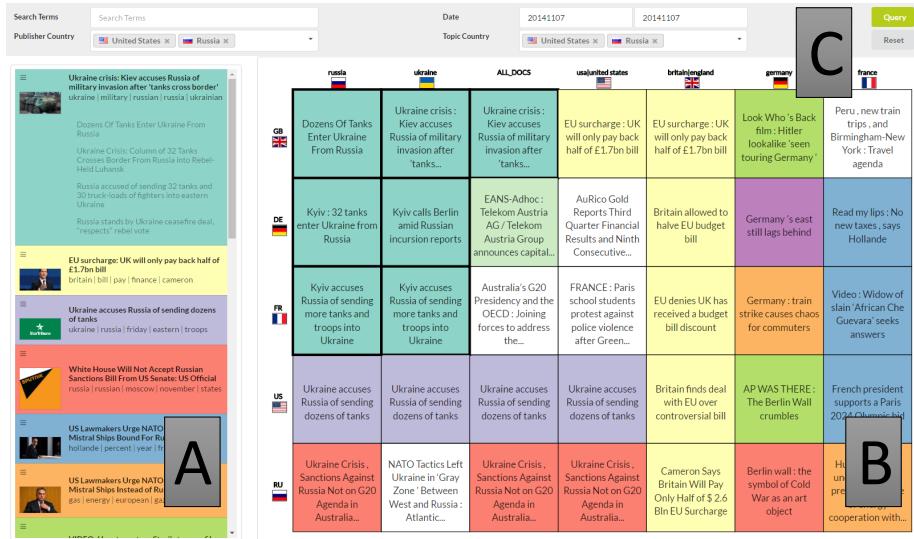

Fig. 6. Main user interface of NewsBird.

The two major components of NewsBird's user interface are the topic list (A) in Fig. 6, and the matrix view (B). The topic list aids the user in quickly reviewing extracted topics. If users find a topic of interest, they can use the matrix view to comparatively explore the cells that are relevant to that topic. By default, each cell in the matrix view shows the summary of the cell's main topic. The components bar (C) enables the user to refine the analysis scope or to change visualization parameters.

To get an overview of the news landscape user can explore the topics in the topic list or the cells in the matrix view. The topics in the list are ordered by importance, i.e., by default according to the number of cells in which the topics occur. Similar to established

systems, NewsBird shows a summary for each topic. We use the summary of the title extracted during the summarization task. Below the title, we show an excerpt of the lead paragraph, and a thumbnail. Users can also view the summaries of related articles.

The matrix is the main component of the visualization and allows users to comparatively explore perspectives on topics of interest. To quickly get an overview and map topics from the topic list to the matrix, each cell's background color matches the color of the cell's main topic in the topic list. Each cell displays the highest scoring summary of its main topic, by default, selected from the title value. If users hover over a cell, a popup window shows more information, such as the lead paragraph and a picture of the cell's main topic. Fig. 7 shows an excerpt of the matrix view for news coverage from November 7, 2014.

Fig. 7. An excerpt of the matrix view showing news coverage from November 7, 2014.

The matrix view and the topic list are visually and interactively connected. For instance, if the user hovers over a cell in the matrix, only the cell's topics will be shown in the topic list. If a user selects a topic in the topic list, NewsBird will re-sort the matrix so that cells containing this topic are placed nearby as shown in Fig. 8.
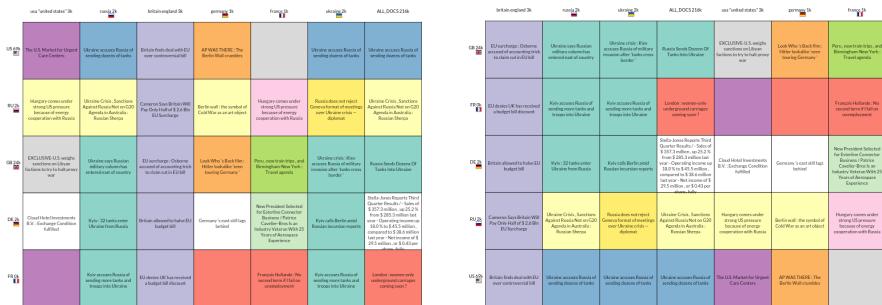

Fig. 8. Comparison of an unsorted matrix (left) and a matrix, which has been sorted so that cells of the selected topic and other major topics are adjacent (right). The topic-to-color mapping enables users to quickly get an overview of extracted topics if it is combined with sorting (right).

The sorted matrix shown in Fig. 8 facilitates recognizing individual topics and allows to efficiently compare the summaries of cells with related topics. If users seek more

information on a specific topic, they can interact with the matrix and the other views, through zooming, panning, and sorting rows and columns of the matrix. Additionally, they can gain more insights on a topic by refining the custom query (see Section 4.2) using the components bar (C) in Fig. 6. Furthermore, the components bar allows users to refine visual parameters, e.g., adjusting the size of cells and colors in the matrix. By default, NewsBird uses colors from *ColorBrewer*, a tool that helps to select appropriate color palettes for visualizations [29]. NewsBird's interactive and (if required) iterative exploration workflow supports deductive and inductive approaches to news analyses.

4.7   Comparison with Established News Aggregators

In this section, we conceptually compare NewsBird with the established news aggregators Google News, Bing News, and Yahoo News regarding the systems' support for bias-aware news analysis.

All established news aggregators employ a list-based overview visualization of current news topics. Due to its ubiquitous use in popular news aggregators, this interface concept is more familiar to users than NewsBird's matrix view. However, the established news aggregators do not offer the same convenient and efficient news overview for custom queries, i.e., if users seek to explore articles for a specific topic or time-period. In such cases, the existing news aggregators only show a list of articles that match the user's query as depicted in Fig. 9. On the contrary, NewsBird allows users to explore the result sets of custom queries using the interactive topic list and matrix view.

Another weakness of established news aggregators is their inability to analyze news from the past. Yahoo News, for instance, only allows the analysis of news from the last week, Bing News does not show news older than a month. Furthermore, Bing News restricts the user to viewing news published in the country and language specified by the user's IP. Google News allows searching articles from the past, but shows them as search results in a basic list depicted in Fig. 9. In contrast, NewsBird does not restrict the set of articles that user can analyze in any way. Users can freely choose the time range and country of origin for articles they want to explore.

Established news aggregators offer a better response time for user queries than NewsBird, because they perform all time-consuming tasks of the analysis workflow (see Section 2.2) including grouping and summarization in regular intervals. The systems then show the precomputed results to users visiting their start page. Our system needs 41 seconds to analyze the 33,000 articles contained in the dataset.

The most significant disadvantage we see in established news aggregators is that none of them focuses on broadening the users' news understanding. Google News, Yahoo News, and Bing News show multiple related articles from different publishers for each news topic (similar to Fig. 5). However, the related articles are neither selected, nor displayed in a way that broaden the users' news understanding. For international news, Google News recently started to also show articles from other countries, but only in the news overview shown on the start page, not for custom user queries. Neither Google, Bing, nor Yahoo have disclosed information on the internal workflows to select related articles. However, we found that the articles suggested for a topic are typically from major publishers, e.g., Reuters and CNBC. The visualization of current news

aggregators will only broaden the user's news understanding if, by chance, these major publishers present different perspectives on the topic.
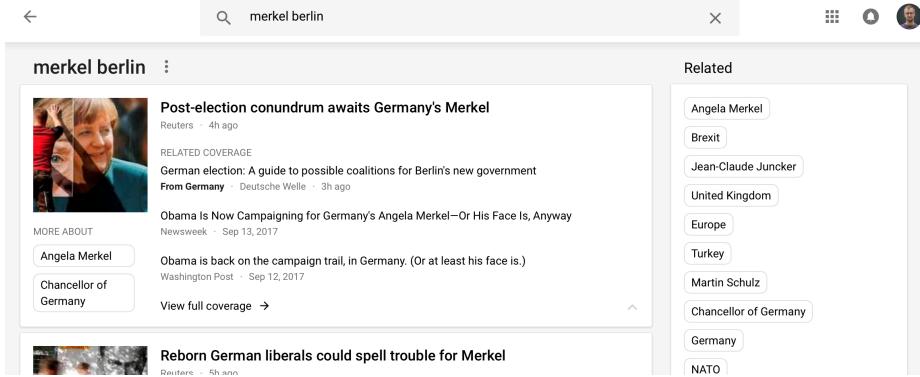

Fig. 9. Established news aggregators only show a list of articles if users issue custom queries.

5   Evaluation

To evaluate the capability of MNA and the NewsBird system to support users in bias-aware news analysis, we conducted a user study. The study compares the list-based visualization of topics and topic-specific articles common to all established news aggregators, such as Google News (cf. Section 2.2), to NewsBird's visualization, which combines the list-based topic visualization with the matrix view. The study addresses the following three research questions.

How do the two visualization approaches affect the users' effectiveness and efficiency in completing the three main tasks in bias-aware news analysis:

RQ1: *getting an overview of news topics on a specific day?*
RQ2: *retrieving details on a specific news topic*?
RQ3: *identifying different perspectives in news coverage on the same topic*?

5.1   Method

Participants: We recruited 10 undergraduate and graduate students (3 female) in the 20-30 age range (mean: 26.4 years), from the humanities (5), computer science (4), and biochemistry (1). We randomly selected five participants to use the matrix-based visualization (group A) and five to use the established list-based visualization (group B).

Scenarios and Tasks: We presented participants with two news scenarios, i.e., news coverage from two specific days. The news articles for both days originate from Europe Media Monitor data. The first news scenario (N1) includes news articles from November 7, 2014 when Russian tanks invaded Ukraine (cf. Table 1). The second news scenario (N2) represents news articles from June 7, 2015 the starting day of the G7 summit in Germany. Also, events related to the Ukraine conflict and the Greek debt crisis made headlines on June 7, 2015. We chose news coverage on past events for two reasons.

First, media bias is easier to identify in retrospective. Second, users are more likely to have in-depth knowledge about the news coverage on current events, which may skew time measurements, which we use as performance metrics.

For each of the two news scenarios, we asked participants to complete three tasks. First, participants should write down headlines that represented the five topics they encountered most frequently in the news coverage of that day. Second, the participants should find and write down a detail related to one of the most important topics on that day. We used Wikipedia[3] to obtain an independently curated list of important events and their details for both days. Third, users should find and write down perspectives that portray the same event differently, e.g., contradicting headlines.

Measurements: We collected the following qualitative and quantitative measurements similar to the study of Park et al. [51].

Before the experiment, we asked the participants to answer five questions about their news consumption. i) *How do you consume news? (E.g., TV, radio, printed newspaper, news websites, online news aggregators.)* ii) [Short explanation of the term news aggregators] *Which news aggregation services are you familiar with?* iii) *Which news aggregation service do you use most often?* iv) *On average, how many hours per month do you spend on your primary news aggregation service?* v) *On average, how many hours per month do spend on consuming news in general?*

After completing each of the two news scenarios, we asked participants to rate the degree of difference they observed for news articles in the scenario on a 6-point Likert scale: "*How do you rate the differences in reporting for the events that exhibited the strongest differences in your view? (1 - minor differences, e.g., different wording is used to report on the same content and presenting the same perspective; 6 - articles differ completely in content or perspective)*".

After completing all three tasks in both news scenarios, we asked participants to rate the effectiveness and efficiency of the visualization they used for each of the three tasks using a 6-point Likert scale: i) *"I am satisfied with the ease of completing task t in scenario s."* ii) *"I am satisfied with the amount of time it took to complete task t in scenario s. (1 - strongly disagree; 6 - strongly agree)"*. With t being 1, 2, or 3 and s being 1 or 2 respectively. Additionally, we asked participants to rate their impression of the system they used regarding the following dimensions using 6-point Likert scales.: i) "*How important do you consider the ability of a system to let the user view and compare different perspectives on the same topic in news coverage? (1 - not important; 6 - very important)*" ii) "*To which degree did the system motivate you to read different articles on the same event or topic? (1 - not at all, 6 - very much)*" iii) "*To which degree did the system show differences in articles on the same event or topic without the need to view the articles themselves? (1 - it did not, 6 - it clearly showed such differences)*" iv) *Does the system help to gain a more balanced and objective understanding of news? (1 - it does not; 6 - it helps a lot)*".

We concluded the experiments by conducting a semi-structured interview with open questions on participants' answers to previous questions and their perception of the

---

[3]   https://en.wikipedia.org/wiki/Portal:Current_events/2014_November_7
     https://en.wikipedia.org/wiki/Portal:Current_events/2015_June_7

system. After the experiment, the two researchers who conducted the experiment independently of each other assessed the nature and the extent of differences that participants found in news articles as part of task three in both scenarios. We rated articles that participants retrieved using the following four-point Likert scale: 0 – articles on different events; 1 – articles that present identical information with the same sentiment in slightly different wording; 2 – articles that present different information with same sentiment; 3 – articles that present opposing information. We averaged the researchers' ratings per participant.

In addition to these qualitative measurements, we recorded the time required by the participants to complete individual tasks in the two news scenarios.

Procedure: We began each experiment by asking participants to complete a questionnaire that gathered background data, e.g., age, field of study, and news consumption. After that, all participants received a 4-minute-long introduction into the functions and the use of the test system, which provided either the list-based or the NewsBird visualization. We read the introduction from a script to ensure identical conditions for all participants. We then presented the participants with the first news scenario and asked them to complete the three tasks. To ensure that all participants of a group saw identical information and could find instances of media bias, we manually selected the dates and analysis scope for both scenarios. After completing all three tasks for N1, we repeated the procedure for N2. Participants could ask questions throughout the experiment. We recorded the screen seen by the participants and the participants' verbal statements for later analysis. The study was not time restricted. The minimum time required for completing the study was 0.75h, the mean 1.14h, and the maximum 1.25h.

5.2 Results

News Consumption: On average, participants spent 26h per month consuming news (minimum 3h, maximum 80h). While seven participants knew at least one news aggregator, most commonly Google News, only one participant stated to use a news aggregator regularly. Eight participants read online news outlets, four read news on social media, three read news in newspapers and magazines, one watched news on YouTube, and one listened to news on the radio. Prior to the experiment, seven participants stated that "the ability of a system to let the user view and compare different perspectives on the same topic in news coverage" is "very important" (6 on the Likert scale). After the experiment, all participants chose this answer.

Despite the importance participants assigned to comparing different news perspective, during the semi-structured interview, five participants stated they have never compared related articles of different news outlets, even if they felt that the coverage on a specific topic was slanted towards one side. Three participants indicated they do so fewer than five times per week. One participant stated that manual comparison was "too tedious", another that he only reads news articles from his favorite publishers, which were all mainstream media of his country, and thus "likely contain identical information anyways". We also asked for which news categories the participants compared or would compare articles. The categories that participants mentioned most frequently were politics (8 participants), tech (3 participants), and economy (2 participants).

Efficiency gains, especially for the task of getting an overview of important topics from multiple news outlets, were the main reason why participants would use a news aggregator. Specifically, four participants stated they would use a ness aggregator for not having to visit multiple news outlets' websites, two would use a news aggregator to avoid reading related articles that contain almost identical information.

Task Performance: In task one of both news scenarios, participants should write down headlines of the five most prevalent news topics of that day. Fig. 10a shows the mean completion times for all three tasks in both news scenarios.
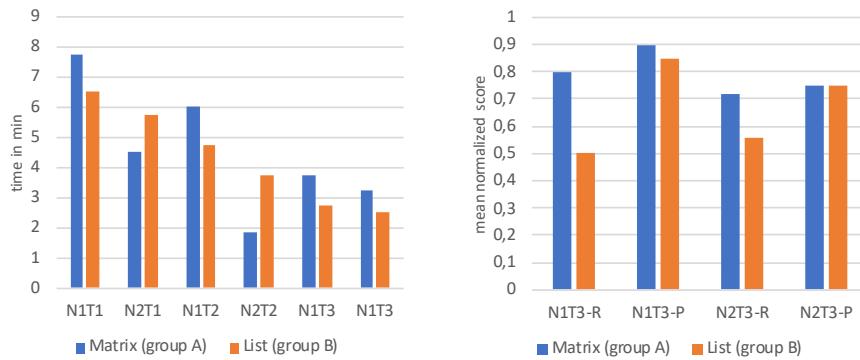

Fig. 10. (a) Mean task completion time in minutes for news scenario N1/2 and task T1-3. (b) Mean normalized ratings of participants (appendix -P) and researchers (appendix -R) on the degree to which information differed in articles retrieved for news scenario N1/2 and task T1-3.

Overall, we found no significant difference in the performance of group A (matrix-based vis.) and group B (list-based vis.) for task one. All participants successfully completed task one. The mean completion time for the task, when taking the average for both news scenarios, was 6.1 minutes in both groups. However, the group that used the matrix-based visualization showed a reduction of 42% in the time required to complete task one in the second news scenario (N2T1) compared to completing task one in the first news scenario (N1T1). The observation is mainly due to learning effects. Four participants in group A stated that they had to get used to the matrix visualization during N1T1; three experimented with the matrix for a while before starting to compete N1T1.
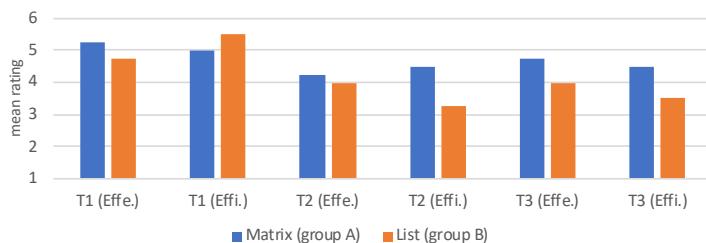

Fig. 11. Participants' ratings on the perceived effectiveness (Effe.) and efficiency (Effi.) for all tasks T1-3 averaged over both news scenarios.

The participants perceived the matrix-based visualization as being slightly more effective, and the list-based visualization to be slightly more efficient for solving task one (cf. Fig. 11). Most participants addressed the task by reviewing the topic list in either of the tested visualizations. By default, the topic lists are sorted by topic frequency. Four participants in group A stated that they used both the topic list and the matrix view to complete N1T1 and N2T1, which could be the reason for the slightly higher effectiveness rating. One participant in group A said that the number of cells highlighted in the same color, enabled him to visually perceive the importance of respective topics, making the first task easier to solve.

Regarding RQ1, these findings indicate that NewsBird's visualization is as effective and efficient as the list-based visualizations employed by established news aggregators for getting an overview of news topics on a specific day.

In task two of both news scenarios, the participants should find a detail related to one of the five most prevalent news topics of that day. As for task one, we did not observe a significant difference in the performance of participants from group A and group B for completing task two. All participants successfully solved the task in both news scenarios. On overage, group A (matrix-based vis.) completed the task slightly faster (4 minutes) than group B (4.3 minutes), cf. Fig. 10a. Similar to task one, we observed a strong decrease (68%) in the completion times for task two in the second news scenario compared to task two in the first news scenario for the group that used the matrix-based visualization. We attribute the difference to the same learning effects described for task one. While we expected participants to solve the task by issuing a custom query in the analysis scope (see Section 4.2), 70% of the participants in both groups initially tried to find the details we asked for by using another component of the user interface. Participants in group A (matrix-based vis.) tried to find the requested information in the matrix view. In both groups, 80% of the participants opened and read articles on the websites of news outlets or checked the article list of related topics. The different approaches to completing the task affected the ratings on perceived efficiency in both groups. Participants of group A had a higher chance to find detail information in the matrix, which natively shows more information than the topic list. Consequently, participants in group A perceived the matrix-based visualization as more efficient (mean rating 4.5 compared to 3.3 for the list-based visualization) for completing task two in both news scenarios (cf. Fig. 11).

Regarding RQ2, these findings indicate that both visualizations offer a comparable effectiveness and efficiency for retrieving additional details on a specific news topic.

In task three of both news scenarios, participants should find different perspectives in news coverage on the same topic. The mean completion time of participants in group A (matrix-based vis.) was slightly higher (3.5 minutes) than the completion time of participants in group B (2.6 minutes) for this task (cf. Fig. 10a). However, participants of group A rated both the effectiveness and the efficiency of completing the task using the matrix-based visualization as significantly higher than the participants in group B (cf. Fig. 11). Participants in group A also retrieved articles that contained a higher amount of differing information than the articles retrieved by participants of group B. Fig. 10b shows the mean normalized ratings of participants (appendix -P) and researchers (appendix -R) on the degree to which information differed in the articles retrieved

for task three in both news scenarios. In all cases, the participants assessed the differences they found in news coverage as more profound than the researchers. However, for group A (matrix-based vis.) the difference in the ratings assigned by participants and researchers was significantly smaller than for group B. Also, participants in group A found more contrastive perspectives (mean normalized score for both news scenarios: 0.76) than participants in group B, who used the list-based visualization (mean normalized score: 0.53). Participants in group B mostly found sub-events, rather than different perspective on the same event. Yet, the participants in group B assessed these information as representing different, almost contradicting perspective. This result indicates that participants who used the list-based visualization did not become aware of the truly different perspectives that exist in news articles of that day. The design of the list-based visualization approach is not intended to show contrastive information or enable bias-aware news analysis (see also Section 2.2)

Regarding RQ3, these findings suggest that NewsBird's visualization is more effective than the list-based visualization employed in established news aggregators in making participants aware of different perspectives in news coverage on the same topic.

Perception of the Systems: Participants of group A who used NewsBird agreed more strongly than participants of group B on the statements that the system:

- motivates them to read different articles on the same event (mean rating group A: 4.8, mean rating group B: 3.3)
- shows differences in news coverage more effectively without the need to view the articles (mean rating group A: 5.3, mean rating group B: 4.5
- helps them to gain a more balanced and objective understanding of news (mean rating group A: 4.8, mean rating group B: 4).

The large majority (90%) of participants on both groups perceived core news aggregation functionality, e.g., viewing important topics and related articles, as helpful. Four participants in group A found the bias-aware features of NewsBird, specifically that the system showed articles from different countries and in different languages (translated to English) as useful. Participants in both groups criticized the topic quality (70%) and suggested to reduce the information load (50%), e.g., because large matrices contained many summaries.

6 Discussion and Future Work

The user study presented in Section 5 supports the findings of our conceptual comparison of NewsBird and established news aggregators (see Section 4.7). The study investigated the three main use cases that we see for bias-aware news analysis: getting a news overview, retrieving details for a specific topic, and broadening the understanding of a topic by exploring different perspectives. The task of giving an overview about important current topics is common to all news aggregators. To perform this task, NewsBird supports the user with a suitable topic selection and relevant details on demand. In contrast to the established news aggregators, the implementation of our MNA concept supports broader news understanding by explicitly showing differences in news

coverage on international news topics. Topic control further enriches the analytical capabilities of NewsBird by letting users narrow down the results to a chosen topic, again organizing the information in countries for better comparison. Users also appreciated that NewsBird shows the matrix view and topic view for custom queries, whereas established news aggregators show the news overview visualization only for the current day. For custom queries the systems show a list of likely relevant articles.

Our plans for future work are twofold. First, we seek to realize several technical improvements of the analysis workflow. Second, we will investigate how the MNA approach can be generalized to enable bias-aware news analysis for additional news categories as well as the application of the MNA concept for use cases beyond news analysis. Hereafter, we describe details on both directions of our future research.

We plan technical improvements for all steps of the analysis workflow, i.e., data gathering and article extraction, matrix initialization, grouping, summarization, and visualization (cf. Fig. 3). To enable MNA for more articles than those in our current, static dataset, we plan to integrate the news-please crawler and extractor for news articles [27] into the data gathering and extraction step. news-please can monitor a website's RSS feed and automatically extract the most recent articles. This addition will enable users of NewsBird to also explore related articles of current events.

NewsBird's current matrix initialization depends on the ability to correctly identify the country mentioned in a news article. While humans can easily derive this information, it is nontrivial to obtain via NLP. Our naive method achieves good results, but cannot handle synonyms and semantically similar phrases, such as "the US" and "The White House". We tested basic query expansion techniques using DBpedia and WordNet to improve the retrieval of relevant articles, but achieved mixed results. We plan to investigate the suitability of other knowledge graphs, such as Yago3 [39], to improve the query expansion step.

Topic extraction is a crucial component for the grouping step of the MNA workflow and was the feature that participants of our user study most frequently requested to improve. We noticed that cells that contain only a few documents in matrices with more than 40 cells will not have their own topic, but rather a (non-fitting) topic from other cells with more documents. Increasing the desired number of topics per cell k_t could help addressing this issue but also yields more artificial topics, i.e., topics that do not represent an actual event. Therefore, we also want to investigate whether different LDA configurations, such as changing the number of cell documents and document length, can achieve better results. Further issues of LDA include mixed descriptiveness, distinctiveness, and reproducibility. Descriptiveness refers to how well an extracted topic represents an event common to a group of articles. Distinctiveness represents the degree to which different topics overlap (see Section 4.4). Reproducibility refers to the stability of the extracted topics across multiple LDA runs. Since LDA is a probabilistic topic extraction approach, results will differ from run to run. We attempted to merge semantically overlapping topics by measuring the fraction of top-ranked LDA terms occurring in different topics. Due to mixed results, we did not enable the topic matching step by default. In the future, we plan to investigate other approaches to match similar topics. One idea is to investigate the LDA document distributions. If two topics have similar topic-document assignments, the topics are potential candidates for merging.

To improve topic extraction, we also investigate approaches that employ external knowledge bases to determine the semantic content of texts. Explicit Semantic Analysis (ESA) [18] is a promising approach that models the semantics of a text by representing the text as a vector in a high-dimensional vector space of semantic concepts. Semantic concepts are topics that are explicitly encoded in a knowledge base corpus. For example, Wikipedia articles are manually curated content covering well-defined topics from a broad spectrum, which is why Wikipedia is commonly used as a knowledge base. ESA allows to include external, human-encoded knowledge into the topic extraction process, whereas LDA, exclusively operates on the given collection.

We expect that NewsBird's grouping accuracy can be improved by devising an automated clustering method that exploits news-specific characteristics. One idea is to extract the so called journalistic five W and one H questions (5W1H) and use them as input features to the clustering process. Journalists usually answer the 5W1H, i.e., who did what, when, where, why, and how, within the first few sentences of a news article [23, 24]. For ESA-based topic modelling, we plan to investigate Semantic Concept Pattern Analysis [42] to group related articles. The approach performs ESA for text segments, e.g., paragraphs, and then analyzes whether semantically similar content is presented in (partially) similar order within articles. Sequential patterns of semantically similar content indicate a structural similarity and thus a higher relatedness of articles. Analyzing the similarity of images in articles in addition to the topics extracted using LDA or ESA is another idea we plan to investigate to refine the grouping of related articles. Images are an important device to frame an article, i.e., influence its perception by the reader. In academic plagiarism detection, analyzing similar images proved valuable for determining semantically similar documents regardless of textual similarity [41]. We expect that images can serve the same goal for news articles, for which NLP methods often fail to determine the subtle differences in reporting as discussed in Section 2.4.

Our summarization method achieves good results, but does not fully exploit the potential of MNA. In the future, we will investigate how we can use ideas from multi-document summarization for *matrix-based summarization*, e.g., by reducing inter-cell redundancy, which could further broaden a user's news understanding. We also found that larger matrices currently contain too much textual information so that users cannot quickly get an overview of the topics present in the analyzed news coverage. To improve the visual analysis of very large matrices, we plan to devise summarization methods that extract tokens or phrases rather than full sentences. Another idea is to visually support the exploration by showing glyphs (instead of summaries) that represent topics and differences when the user views a very large matrix. We expect that the resulting visual patterns would help to view commonalities and differences at large scale.

Currently, NewsBird employs a matrix to visualize the analysis results, since MNA is built around structuring and analyzing articles and topics in a matrix. Our evaluation has shown that this rather simple visualization already enables users to become aware of different information on the same topic. However, we also want to examine other visualizations to improve the usability of the system. For example, for coverage on international news topics that likely differs between countries, a geo-based visualization may achieve better results.

Our vision is to make bias-aware news analysis available to the public. While MNA is a general analysis approach to reduce media bias, NewsBird currently only supports the exploration of international news topics. In our view, regular news consumers will benefit the most from MNA if established news aggregators visualize topics similarly as suggested in Fig. 5. To achieve this goal, our future research will focus on the generalization of the matrix initialization step to enable bias-aware news-analysis for additional news categories.

A first task will be to develop dimensions that enable the comparative analysis of more than international news topics. For instance, we could divide articles using predefined groups of news sources, e.g., sources that often publish the same perspective or sources that use similar slant. Such information could also be computed automatically. A basic approach could be to simply divide articles by publisher, e.g., if the user wants to get an overview of how specific publishers portray an event. This approach could help to reveal different perspectives for news topics that only a few articles report on, e.g., local politics.

We also need to investigate which dimensions are suitable to split related news articles into diverse subgroups. We seek to find additional properties in news articles that indicate which dimensions will maximize the expected diversity of cells in the matrix for a given news situation. Approaches to achieve this goal could be to ask users to set contrary sources in relation [52], or to identify text characteristics in different articles and subsequently comparing such differences on article and sub-article level [51].

Lastly, we want to investigate more use cases for MNA. In our view, the MNA concept can be applied to any data that expresses a relation between two entities, such as emails, e.g., the *Enron Email Dataset* [33], and other messages. MNA could also be used to analyze product reviews, e.g., by spanning a matrix over publishers and products or features of one product.

7 Conclusion

This article describes *matrix-based news analysis* (MNA), a news exploration approach that supports users in discovering different perspectives in news coverage to reduce the effects of media bias. MNA organizes news content according to inherent attributes, e.g., publisher and mentioned country, and visualizes this information to reveal diverse and controversial information that can hardly be found using established news aggregators. For this purpose, MNA organizes news articles in the cells of a matrix spanned over two dimensions, which are selected to maximize the expected diversity in the resulting cells. For instance, the analysis of what is stated in one country about another country can help to understand international news topics while reducing media bias by revealing different perspectives of the (involved) countries.

We also present *NewsBird*, an extensible news aggregator that implements MNA to explore international news topics. We evaluated the capabilities of the NewsBird system with the help of a user study. We found that NewsBird achieves similar effectiveness and efficiency as the list-based visualization approach employed by all established news aggregators, such as Google News, for the two primary use cases of regular news

consumption, i.e., (1) getting an overview of important topics, and (2) retrieving further details for specific topics upon users' demand. More importantly, we found that users of NewsBird were significantly more effective than users provided with the list-based visualization of established news aggregators in becoming aware of different perspective on the same news topics, which is the essential goal of bias-aware news analysis.

The first implementation of NewsBird already achieves promising results, which motivate us to continue our research on the identification of media bias. Our primary direction for future research will be to improve the analysis workflow and to generalize NewsBird to support additional news categories and other use cases.

NewsBird is an open-source system and its code is available under the Apache 2 license at: https://github.com/fhamborg/NewsBirdServer

Acknowledgments

References

[1]   Alexander, E. and Gleicher, M. 2016. Task-Driven Comparison of Topic Models. *IEEE Transactions on Visualization and Computer Graphics*. 22, 1 (2016), 320–329.

[2]   Atkinson, M. and der Goot, E. 2009. Near real time information mining in multilingual news. *Proceedings of the 18th international conf. on World wide web* (2009), 1153–1154.

[3]   Baker, B.H. et al. 1994. *How to identify, expose & correct liberal media bias*. Media Research Center Alexandria, VA.

[4]   Balahur, A. et al. 2013. Sentiment analysis in the news. *arXiv preprint:1309.6202*. (2013).

[5]   Baron, D.P. 2006. Persistent media bias. *Journal of Public Economics*. 90, 1 (2006), 1–36.

[6]   Bernhardt, D. et al. 2008. Political polarization and the electoral effects of media bias. *Journal of Public Economics*. 92, 5 (2008), 1092–1104.

[7]   Besley, T.J. and Prat, A. 2002. Handcuffs for the grabbing hand? Media capture and government accountability. (2002).

[8]   Best, C. et al. 2005. *Europe media monitor*.

[9]   Blei, D.M. 2012. Probabilistic topic models. *Comm. of the ACM*. 55, 4 (2012), 77–84.

[10] Bui, C. 2010. How online gatekeepers guard our view: News portals' inclusion and ranking of media and events. *Global Media J.* 9, 16 (2010), 1–41.

[11] Business Insider 2014. These 6 Corporations Control 90% Of The Media In America.

[12] Cambria, E. and White, B. 2014. Jumping NLP curves: a review of natural language processing research [review article]. *Comp. Intelligence Mag., IEEE*. 9, 2 (2014), 48–57.

[13] Cuthill, E. and McKee, J. 1969. Reducing the Bandwidth of Sparse Symmetric Matrices. *Proceedings of the 1969 24th national conference of the ACM*. (1969), 157–172.

[14] Domke, D. et al. 1999. The politics of conservative elites and the "liberal media" argument. *Journal of Communication*. 49, 4 (1999), 35–58.

[15] Esser, F. et al. 2001. Spin Doctors in the United States, Great Britain, and Germany

Metacommunication about Media Manipulation. *Int. J. Press/Politics*. 6, 1 (2001), 16–45.

[16] Evans, D.K. et al. 2004. Columbia newsblaster: multilingual news summarization on the Web. *Demonstration Papers at HLT-NAACL 2004* (2004), 1–4.

[17] Flaounas, I. et al. 2010. The Structure of the EU Mediasphere. *PLoS ONE*. 5, 12 (2010).

[18] Gabrilovich, E. and Markovitch, S. 2005. Feature Generation for Text Categorization Using World Knowledge. *Proc. 19th Int. J. Conf. for Art. Intelligence* (2005), 1048–1053.

[19] GALLUP 2015. Americans' Trust in Media Remains at Historical Low.

[20] Gauthier, G. 1993. In defence of a supposedly outdated notion: The range of application of journalistic objectivity. *Canadian Journal of communication*. 18, 4 (1993), 497.

[21] Gentzkow, M. and Shapiro, J. 2005. *Media bias and reputation*.

[22] Griffiths, T.L. and Steyvers, M. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences*. 101, suppl 1 (2004), 5228–5235.

[23] Hamborg, F. et al. 2018. Extraction of Main Event Descriptors from News Articles by Answering the Journalistic Five W and One H Questions. *Proc. of the ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)* (2018).

[24] Hamborg, F. et al. 2018. Giveme5W: Main Event Retrieval from News Articles by Extraction of the Five Journalistic W Questions. *Proceedings of the iConference 2018* (Sheffield, UK, 2018).

[25] Hamborg, F. et al. 2017. Identification and Analysis of Media Bias in News Articles. *Proceedings of the 15th International Symposium of Information Science* (2017).

[26] Hamborg, F. et al. 2017. Matrix-based News Aggregation: Exploring Different News Perspectives. *Proc. of the ACM/IEEE Joint Conference on Digital Libraries*. 10, 17 (2017).

[27] Hamborg, F. et al. 2017. news-please: A Generic News Crawler and Extractor. *Proceedings of the 15th International Symposium of Information Science* (2017), 218–223.

[28] Harcup, T. and O'neill, D. 2001. What is news? Galtung and Ruge revisited. *Journalism studies*. 2, 2 (2001), 261–280.

[29] Harrower, M. and Brewer, C.A. 2011. ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *The Map Reader*. 261–268.

[30] Hatcher, E. and Gospodnetic, O. 2004. Lucene in action. (2004).

[31] Hu, M. and Liu, B. 2004. Mining and summarizing customer reviews. *Proc. of the 10th ACM SIGKDD int. conference on Knowledge discovery and data mining* (2004), 168–177.

[32] Huang, X. et al. 2011. Comparative news summarization using linear programming. *Proc. of the 49th Ann. Meet. of the ACL: Human Language Tech.: Volume 2* (2011), 648–653.

[33] Klimt, B. and Yang, Y. 2004. The enron corpus: A new dataset for email classification research. *Machine learning: ECML 2004*. Springer. 217–226.

[34] Kohlschütter, C. et al. 2010. Boilerplate detection using shallow text features. *Proc. of the third ACM int. conf. on Web search and data mining* (2010), 441–450.

[35] Krstaji, M. et al. 2011. Cloudlines: Compact display of event episodes in multiple time-series. *IEEE Trans. on Visualization and Computer Graphics*. 17, 12 (2011), 2432–2439.

[36] Krstajic, M. et al. 2013. Story Tracker: Incremental visual text analytics of news story development. *Information Visualization*. 12, 3–4 (2013), 308–323.

[37] Kull, S. et al. 2003. Misperceptions, the media, and the Iraq war. *Political Science Quarterly*. 118, 4 (2003), 569–598.

[38] Lin, C.-Y. and Hovy, E. 2002. From single to multi-document summarization: A prototype system and its evaluation. *Proc. of the 40th Annual Meeting on ACL* (2002), 457–464.

[39] Mahdisoltani, F. et al. 2015. YAGO3: A Knowledge Base from Multilingual Wikipedias. *Proceedings of CIDR*. (2015), 1–11.

[40] McKeown, K.R. et al. 2002. Tracking and summarizing news on a daily basis with Columbia's Newsblaster. *Proc. of the 2nd int. conf. on HLTR* (2002), 280–285.

[41] Meuschke, N. et al. 2018. An Adaptive Image-based Plagiarism Detection Approach. *ACM/IEEE-CS Joint Conf. on Digital Libraries (JCDL)* (2018).

[42] Meuschke, N. et al. 2017. Analyzing Semantic Concept Patterns to Detect Academic Plagiarism. *Proc. of the 6th Int. Work. on Mining Scientific Publications* (2017), 46–53.

[43] Mitchell, R. 2015. *Web scraping with Python: collecting data from the modern web*. O'Reilly Media, Inc.

[44] Mullainathan, S. and Shleifer, A. 2005. The market for news. *American Economic Review*. (2005), 1031–1053.

[45] Munson, S.A. et al. 2013. Encouraging Reading of Diverse Political Viewpoints with a Browser Widget. *ICWSM* (2013).

[46] Munson, S.A. et al. 2009. Sidelines: An Algorithm for Increasing Diversity in News and Opinion Aggregators. *ICWSM* (2009).

[47] Newman, N. et al. 2015. Reuters Institute Digital News Report 2015. (2015).

[48] News Bias Explored - The art of reading the news: 2014. *http://umich.edu/newsbias/*. Accessed: 2017-08-22.

[49] Oelke, D. et al. 2012. Visual Analysis of Explicit Opinion and News Bias in German Soccer Articles. (2012).

[50] Paliouras, G. et al. 2008. PNS: A personalized news aggregator on the web. *Intelligent interactive systems in knowledge-based environments*. Springer. 175–197.

[51] Park, S. et al. 2009. NewsCube: Delivering Multiple Aspects of News to Mitigate Media Bias. *Proceedings of CHI '09, the SIGCHI Conf. on Human Factors in CS*. (2009).

[52] Park, S. et al. 2011. NewsCube 2.0: An Exploratory Design of a Social News Website for Media Bias Mitigation. *Workshop on Social Recommender Systems* (2011).

[53] Park, S. et al. 2011. The politics of comments: predicting political orientation of news stories with commenters' sentiment patterns. *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (2011), 113–122.

[54] Radev, D.R. et al. 2000. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. *Proceedings of the 2000 NAACL-ANLP Workshop on Automatic summarization* (2000), 21–30.

[55] Recasens, M. et al. 2013. Linguistic Models for Analyzing and Detecting Biased Language. *Proc. of the 51st Ann. Meet. on Association for Computational Ling*. (2013), 1650–1659.

[56] Salton, G. et al. 1975. A vector space model for automatic indexing. *Communications of the ACM*. 18, 11 (1975), 613–620.

[57] Smith, A. et al. 2014. Hiérarchie: Interactive Visualization for Hierarchical Topic Models. *Proc. of the Work. on Interactive Lang. Learn., Vis., and Interfaces*. (2014), 71–78.

[58] Straubhaar, J.D. 2000. *Media Now: Comm. Media in Information Age*. Thomson Learning.

[59] Sundar, S.S. 1999. Exploring receivers' criteria for perception of print and online news. *Journalism & Mass Comm. Quarterly*. 76, 2 (1999), 373–386.

[60] Weskamp, M. 2016. newsmap.

[61] 2014. Moscow to Kiev: Stick to Minsk ceasefire, stop making false "invasion" claims. *RT*.

[62] 2014. Tank column crosses from Russia into Ukraine: Kiev military. *CNBC*.