Matrix-based News Aggregation: Exploring Different News Perspectives

Felix Hamborg
Computer and Information Science
University of Konstanz
felix.hamborg@uni-konstanz.de

Norman Meuschke
Computer and Information Science
University of Konstanz
norman.meuschke@uni-konstanz.de

Bela Gipp
Computer and Information Science
University of Konstanz
bela.gipp@uni-konstanz.de

ABSTRACT

News aggregators capably handle the large amount of news that is published nowadays. However, these systems focus on identifying and presenting important, common information in news, but do not reveal different perspectives on the same topic. Differences in the content or presentation of news are referred to as media bias, which can have severe negative effects. Given their analysis approach, current news aggregators cannot effectively reveal media bias. To address this problem, we present *matrix-based news analysis* (MNA), 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*, a news aggregator that implements MNA for international news topics. The results of a case study demonstrate that NewsBird broadens the user's news understanding while providing similar news aggregation functionalities as established systems.

CCS CONCEPTS

•**Information systems → Web searching and information discovery;** *Information extraction;* Information systems applications; Recommender systems; Summarization; •**Applied computing →** Document management and text processing;

KEYWORDS

Media bias, news aggregation, frame analysis, content analysis, Google News.

ACM Reference format:
Felix Hamborg, Norman Meuschke, and Bela Gipp. 2017. Matrix-based News Aggregation: Exploring Different News Perspectives. In *Proceedings of the 17th ACM/IEEE-CS on Joint Conference on Digital Libraries, Toronto, Ontario, Canada, June 2017 (JCDL'17),* 10 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The coverage of media outlets often exhibits *media bias*, e.g., due to political interference, lobbyism, or ideological focus [49]. 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 [10], which results in a high chance of media manipulation [14, 47]. Trust in media is at a historical low, e.g., less than half of U.S. readers trust media and think it is objective [17].

Table 1: Different headlines for the same event.

Source	Headline
CNBC [12]	Tank column crosses from Russia into Ukraine: Kiev military
RT [44]	Moscow to Kiev: Stick to Minsk ceasefire, stop making false invasion claims

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

Media bias has many, severe effects (cf. [5]). For example, a 2003 survey showed significant differences in the presentation of information on US television channels. Fox News viewers were most misinformed about the Iraq war. Over 40% thought that weapons of mass destruction had been found in Iraq [28] – a false claim the US government used as a justification for the war.

Although more and more information from around the world is available online and often at no cost, many news readers only consult a small subset of news sources [36]. Reasons include the overwhelming number of sources, language barriers, or simply habit. These and further factors can cause a *narrow news perspective* [34], and thus a skewed or incomplete perception of information.

News aggregators, such as Google News, enable 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 *diverse news analysis* aim to reduce media bias by finding different perspectives on the same topic [35, 39, 40]. However, due to limitations of underlying *natural language processing* (NLP) methods, these systems suffer from practical limitations, such as relying heavily on user feedback.

In this paper, we present a novel approach named *matrix-based news analysis* (MNA) that consists of 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 presenting different, possibly biased perspectives on a topic. Our news aggregator *NewsBird* exemplifies MNA for international news topics.

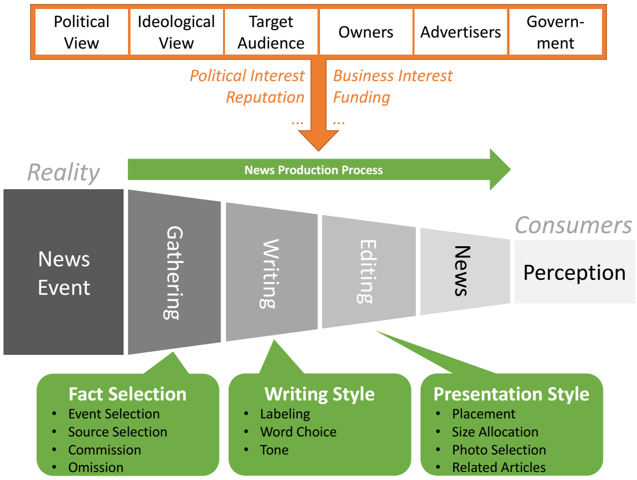


Figure 1: Reasons and forms of media bias. Based on [39].

We structure the remainder of this paper 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.

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 [5] 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 [19, 33], and external factors, such as that journalists emphasize privately obtained information [4], or that governments influence publishers in their favor [6]. 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 [22].

In the social sciences, different perspectives in news coverage are called *frames*. Social scientists use *frame analysis* [31] to inductively find (new) frames in a set of news texts or to deductively find evidence for assumed frames. Both approaches to frame analysis require much manual effort and expertise on the topic.

Figure 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 [39]. During *gathering*, journalists select events, sources, and the facts they want to present. These selection processes bias the resulting news story. During *writing*, journalists can affect the reader's perception of a topic, e.g., through word choice (whether the author uses a positive or negative word to refer to an entity, such as "coalition forces" vs. "invasion forces"), or by varying the credibility ascribed to the source [2, 19, 37].

The third stage, *editing*, determines the (visual) presentation of an article, e.g., through placement (a front-page article receives the most attention). Finally, consumers read the news. Reading may also yield different perceptions of the event [3, 48], but this influence is beyond the focus of this paper.

In conclusion, media bias is a structural, often intentional, flaw inherent to news publishing [13] and can critically impact people's opinions and decisions. Before Section 2.3 describes related work on diverse 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.

2.2 News Aggregation

This section gives a brief introduction to news aggregation, a state-of-the-art approach to let readers overview the large amount of news that is produced nowadays. The analysis pipelines of most news aggregators find the most important news and summarize them for users. This typically involves the following steps, cf. [15]:

(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 [32]. Articles can be extracted using naive approaches, such as website specific wrappers [38], or more generic methods using content heuristics [27]. 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 [21]. 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) [8] as for instance applied in the Europe Media Monitor [7], or clustering methods, such as hierarchical agglomerative clustering (HAC) as used in [30, 39]. 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* [42].

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. E.g., newsmap features a two-level treemap to show news categories and topics [51]. *Hirarchie* shows a topic hierarchy in a sunburst diagram to let users explore different semantics of a topic [46]. 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 perspective are *Newsblaster*, which is one of the first academic news aggregators [30], the *Europe Media Monitor*, which improves automatically aggregated news through manual revision [7], and *PNS*, a news aggregator that provides user personalization [38].

The presented analysis pipeline (see also [27, 36, 7]) enables news aggregators 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 [36, 7] indicate the maturity of the systems and the analysis approach.

However, no news aggregator focuses on revealing differences between related articles [39] 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 [9, 50].

2.3 Diverse News Analysis

This section presents approaches and systems to reduce the effects of media bias – a feature missing in established news aggregators despite being requested by the users of such systems [35].

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 *diverse 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 [39]. 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 phase: NewsCube extracts aspects from each article using keywords and syntactical rules. The systems then weights 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 aspect is considered to be. NewsCube then performs HAC to group related articles. The offered visualization is similar to 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.

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 [40]. 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 [35]. 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 [41].

Comparative or *contrastive news summarization* methods let users view both common and differing sections in groups of related articles (see Section 2.4). However, summarization methods for news texts achieve poor results due to a missing alignment of comparable sections [15] or processing topics instead of articles on a topic [25].

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 [39]. Classic NLP techniques typically rely on statistics and are "[...] just a first step towards natural language understanding" [11]. 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 share many topic-specific keywords.

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 [18]. While sentiment analysis yields good results for texts in which authors explicitly state their opinion, such as product reviews [24], the results for news texts are not satisfactory [37]. 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 aforementioned limitations of NLP methods. For example, comparative summarization methods list the most important common and differing information in multiple news articles [15] or topics [25]. However, the quality of section alignment requires further improvement before these techniques allow for an effective comparison [15]. Another approach uses recursive topic modeling to find different (semantic) components of a topic [46]. The resulting topic components are not always meaningful, but represent artificial subtopics.

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. 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 Discussion

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 the same event. However, 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. Yet, there are approaches that can reduce media bias effects by broadening users' understanding of news topics. These approaches, however, suffer from several practical limitations, such as being restricted to the analysis of one news

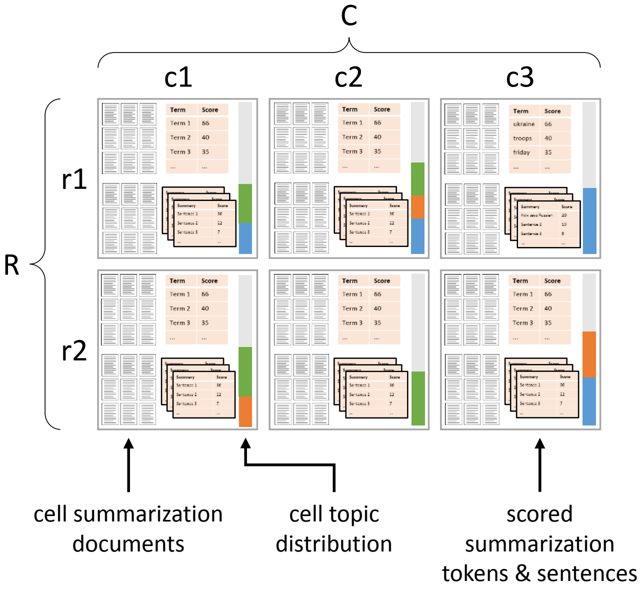


Figure 2: Organization of articles, topics, and summaries.

category [35, 39], requiring manually created knowledge bases [40], and being fine-tuned for specific analysis tasks.

3 MATRIX-BASED NEWS ANALYSIS

In this section, we introduce *matrix-based news analysis* (MNA), 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 phase.

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 whole media landscape of countries or regions. Rows and columns can encode different entity types. In the example shown in Figure 2, the matrix elements show the main content that the media in one country report about another country, i.e. what a reader from one country would typically read about another country.

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).

Figure 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 paper focuses on the description of the novel components of MNA. For the following description, we thus assume that a dataset of news articles exists.

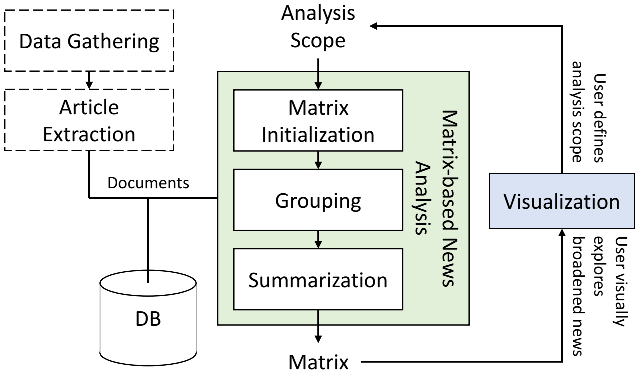


Figure 3: Analysis workflow.

To start the analysis, the user has to define the *analysis scope* (see also Section 4.3), 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.

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) a *topic summary* for each topic: MNA considers all documents containing the topic to create this summary; (2) a *cell summary* for each cell: MNA considers all documents in a cell to create this summary; (3) a *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 Figure 2 – each cell contains one or more weighted topics and the corresponding summaries. Users can control the analysis process, 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 process by incorporating their domain-knowledge.

4 SYSTEM DESCRIPTION: NEWSBIRD

4.1 System Overview

NewsBird is a news aggregator that enables diverse 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 Figure 3. The next Sections describe the capabilities of the system and how we realized them. The description follows the MNA analysis approach described in Section 3.

4.2 Data Gathering and Article Extraction

The current version of NewsBird uses a fixed dataset instead of performing data gathering and article extraction, since the system currently focuses on demonstrating the novel components of the MNA approach. The dataset originates from the Europe Media Monitor [1] and consists of 1.6 million articles gathered from approx. 4,000 publishers from over 100 countries during the period October to November 2014. Articles are available in various languages. We translated all non-English articles to English using a machine translation service[1]. Translating articles was necessary to cope with the limitations of NLP technologies we employ, such as topic modeling, which is significantly less reliable when performed across languages. The quality of machine-translated text is lower than the quality of manually translated text, yet often high enough for IR tasks [15, 16] 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 in an Apache Lucene index. Lucene performs state-of-the-art text preprocessing, such as tokenization, lowercasing, stop word removal, and stemming [23].

4.3 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.

4.4 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. For instance, the cell query of `RU-UA` will only retrieve documents published in Russia that mention Ukraine.

For each cell, the system retrieves 100 cell documents to form a representative sample of all cell documents, while limiting the time required for the subsequent topic extraction and summarization steps. 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.

The two country dimensions enable users to comparatively explore international news topics. The time-range dimension allows users to analyze the development of one or more topics over time, e.g., by showing one day in each column. To support more news categories, additional dimensions can be easily added to the system (see Section 6).

4.5 Topic Extraction to Find Related Articles

NewsBird employs LDA to extract a list of topics from the matrix and assigns documents with identical topics to the same group. To extract topics, NewsBird performs four subtasks.

The first subtask is *text extraction*. We consider the full content (see Section 4.2) of each document, since LDA performs better on 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, where k controls the granularity of resulting topics. In all our tests, we achieved the best results with $k = 2$, as this allows each cell to have one or two cell-specific topics, i.e. topics that semantically represent 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, cf. [20]. 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* the resulting topics and mappings. Theoretically, each cell contains a weighted proportion of all

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

topics – most of them with 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). We do not merge similar topics.

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. This query consists of weighted *topic sub-queries*, one for each of the cell's topics. Each of the topics consists of the topic's 15 highest weighted terms, of which at least 7 must be contained in a document to be retrieved. This parametrization is a trade-off between precision and recall. The resulting documents are the *cell summarization documents*. Users can modify all parameters in the visualization.

4.6 Summarization

NewsBird uses the sum of TF-IDF and topic weights to form the summary score. For each cell, we take all cell summarization documents (shown in Figure 2), and compute summary scores for each of the following three textual dimensions. First, the score of each *token* w is calculated as:

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

where d is a document in a set of documents D. After computing the scores of each token, we calculate 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 [43, 45], we provide the summary score $S(p, D, T)$ that additionally takes into account the sentences' relation to the cell's topics:

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

where $c(t)$ is a function that returns the weight of a topic t in the current cell and $t(w, t)$ returns the term w's weight in topic t. We set the constant $\alpha = 9400$ to balance the unnormalized topic weights and TF-IDF scores.

To improve the quality of summaries, we adjust scores using rules from [29], 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 [39]. 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 previously described cell summaries, NewsBird also summarizes each of the topics itself.

4.7 Visualization

The first step in the exploration process is to define the analysis scope. After the system has completed the analysis defined by the analysis scope, results are visualized as shown in Figure 4. Users can start the news exploration by getting on overview of the news landscape. The *topic list*, component (A) in Figure 4, aids the user in quickly reviewing extracted topics. If users find a topic of interest, they can use the *matrix view* (B) to comparatively explore the individual cells that are relevant to that topic. By default, each cell shows the summary of its main topic. The user can interactively choose other topics from the topic list. NewsBird also enables users to refine the analysis scope (C). For instance, if users seek more detailed information on a topic, they can refine the custom query (see Section 4.3). They can either specify a topic or select a topic from the list (A).

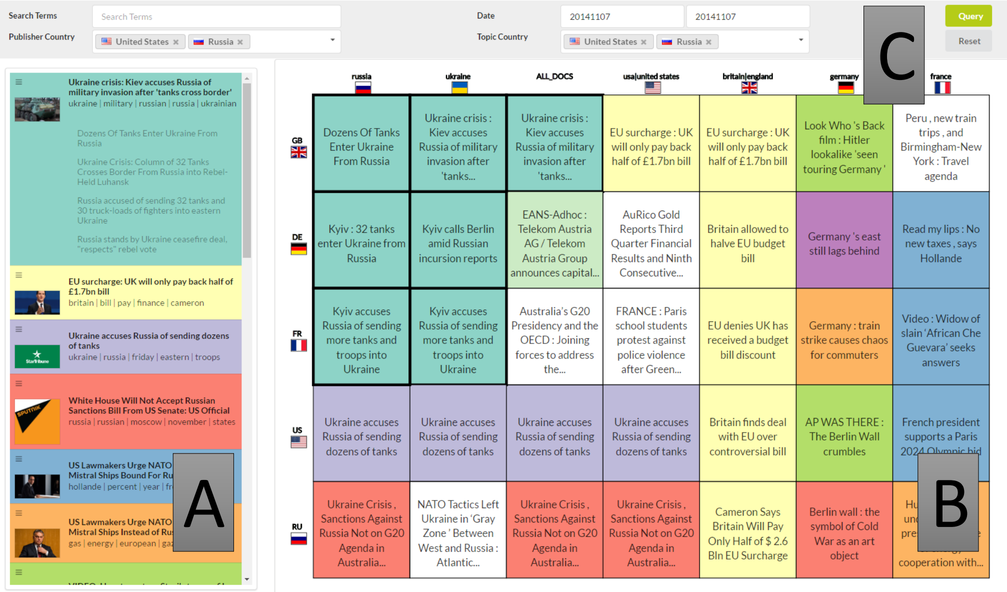
The topics in the list are ordered by their importance, i.e. by default according to the number of cells in which they occur. Similar to established systems, NewsBird shows a summary for each topic. We use the summary of the title extracted during the summarization task. Below the title, we show an excerpt of the lead paragraph, and a thumbnail. Users can also view the summaries of related articles. If the user selects a topic, NewsBird will re-sort the matrix so that cells containing this topic are placed nearby, which enables efficient comparison as shown in Figure 7.

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 windows will show more information, such as the lead paragraph and a picture of the cell's main topic.

The two major components of NewsBird's user interface, the matrix view and the topic list, depicted in Figure 4, 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, all cells in the matrix containing this topic are placed together and display information on the selected topic. The components bar (top) enables the user to open panels that allow to control the query and the visualization.

5 EVALUATION

This section describes a case study with three use cases that demonstrate the capabilities of NewsBird. To illustrate its performance in a real world news consumption context, we conducted the case study with a Google News user. We collected observational and think-aloud-data, and conducted a follow-up interview during which the user assessed the subjective experience while using the system. The dataset for this study contained over 33,000 news articles from November 7th, 2014, covering major news topics of that time, such as the Ukraine crisis and the Brexit discussions. Throughout the study, the user followed the news exploration workflow (see Section 4.7) without being instructed to do so or being aware of it.



Search Terms: Search Terms
Publisher Country: United States, Russia
Date: 20141107 20141107
Topic Country: United States, Russia

Ukraine crisis: Kiev accuses Russia of military invasion after 'tanks cross border'
ukraine | military | russian | russia | ukrainian

Dozens Of Tanks Enter Ukraine From Russia

Ukraine Crisis: Column of 32 Tanks Crosses Border From Russia into Rebel-Held Luhansk

Russia accused of sending 32 tanks and 30 truck-loads of fighters into eastern Ukraine

Russia stands by Ukraine ceasefire deal, "respects" rebel vote

EU surcharge: UK will only pay back half of £1.7bn bill
britain | bill | pay | finance | cameron

Ukraine accuses Russia of sending dozens of tanks
ukraine | russia | friday | eastern | troops

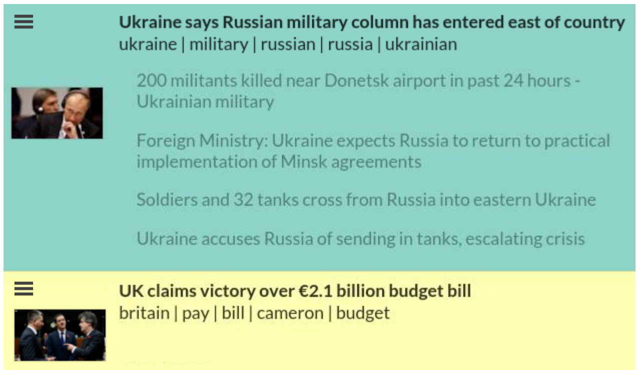
White House Will Not Accept Russian Sanctions Bill From US Senate: US Official
russia | russian | moscow | november | states

US Lawmakers Urge NATO Mistral Ships Bound For Ru
hollande | percent | year | fr

US Lawmakers Urge NATO Mistral Ships Instead of Ru
gas | energy | european | gaz

	russia	ukraine	ALL_DOCS	usa|united states	britain|england	germany	france
GB	Dozens Of Tanks Enter Ukraine From Russia	Ukraine crisis : Kiev accuses Russia of military invasion after 'tanks...	Ukraine crisis : Kiev accuses Russia of military invasion after 'tanks...	EU surcharge : UK will only pay back half of £1.7bn bill	EU surcharge : UK will only pay back half of £1.7bn bill	Look Who 's Back film : Hitler lookalike 'seen touring Germany '	Peru , new train trips , and Birmingham-New York : Travel agenda
DE	Kyiv : 32 tanks enter Ukraine from Russia	Kyiv calls Berlin amid Russian incursion reports	EANS-Adhoc : Telekom Austria AG / Telekom Austria Group announces capital...	AuRico Gold Reports Third Quarter Financial Results and Ninth Consecutive...	Britain allowed to halve EU budget bill	Germany 's east still lags behind	Read my lips : No new taxes , says Hollande
FR	Kyiv accuses Russia of sending more tanks and troops into Ukraine	Kyiv accuses Russia of sending more tanks and troops into Ukraine	Australia's G20 Presidency and the OECD : Joining forces to address the...	FRANCE : Paris school students protest against police violence after Green...	EU denies UK has received a budget bill discount	Germany : train strike causes chaos for commuters	Video : Widow of slain 'African Che Guevara' seeks answers
US	Ukraine accuses Russia of sending dozens of tanks	Ukraine accuses Russia of sending dozens of tanks	Ukraine accuses Russia of sending dozens of tanks	Ukraine accuses Russia of sending dozens of tanks	Britain finds deal with EU over controversial bill	AP WAS THERE : The Berlin Wall crumbles	French president supports a Paris 2024 Olympic bid
RU	Ukraine Crisis , Sanctions Against Russia Not on G20 Agenda in Australia...	NATO Tactics Left Ukraine in 'Gray Zone ' Between West and Russia : Atlantic...	Ukraine Crisis , Sanctions Against Russia Not on G20 Agenda in Australia...	Ukraine Crisis , Sanctions Against Russia Not on G20 Agenda in Australia...	Cameron Says Britain Will Pay Only Half of $ 2.6 Bln EU Surcharge	Berlin wall : the symbol of Cold War as an art object	Hu un pre cooperation with...

Figure 4: Main user interface of NewsBird.



Ukraine says Russian military column has entered east of country
ukraine | military | russian | russia | ukrainian

200 militants killed near Donetsk airport in past 24 hours - Ukrainian military

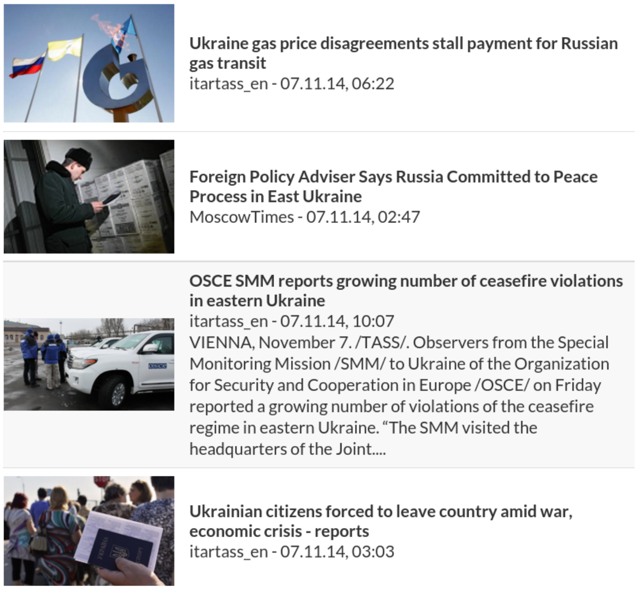
Foreign Ministry: Ukraine expects Russia to return to practical implementation of Minsk agreements

Soldiers and 32 tanks cross from Russia into eastern Ukraine

Ukraine accuses Russia of sending in tanks, escalating crisis

UK claims victory over €2.1 billion budget bill
britain | pay | bill | cameron | budget

Figure 5: Topic list excerpt (first topic is expanded).



Ukraine gas price disagreements stall payment for Russian gas transit
itartass_en - 07.11.14, 06:22

Foreign Policy Adviser Says Russia Committed to Peace Process in East Ukraine
MoscowTimes - 07.11.14, 02:47

OSCE SMM reports growing number of ceasefire violations in eastern Ukraine
itartass_en - 07.11.14, 10:07
VIENNA, November 7. /TASS/. Observers from the Special Monitoring Mission /SMM/ to Ukraine of the Organization for Security and Cooperation in Europe /OSCE/ on Friday reported a growing number of violations of the ceasefire regime in eastern Ukraine. "The SMM visited the headquarters of the Joint....

Ukrainian citizens forced to leave country amid war, economic crisis - reports
itartass_en - 07.11.14, 03:03

Figure 6: Article list excerpt (third article is expanded).

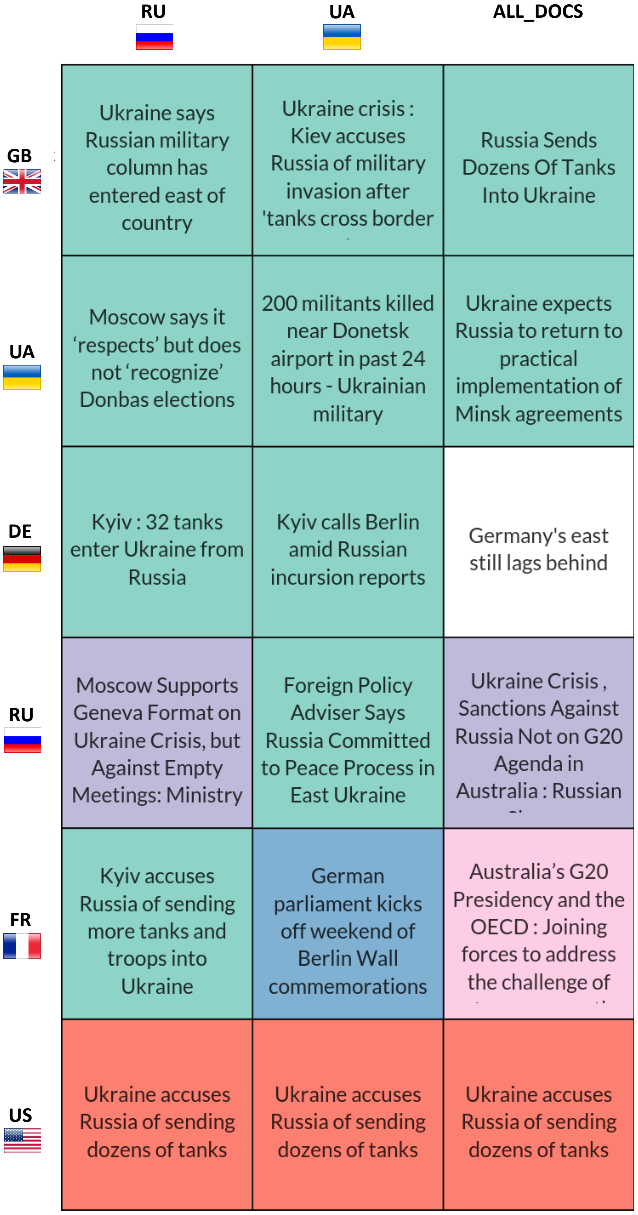
5.1 News Topic Overview

The primary use case of a news aggregator is to provide users with an overview of today's most important news topics. The user in our case study started the interaction with the wish to inform himself about general news and events on November 7th, 2014. First, he scanned the topic list and got interested in the most frequent topic, the Russian tank invasion (first topic in Figure 5). To get additional information, he clicked on the list item to view four supplementary summaries from related articles for this topic.

With the topic list being sorted by importance, it is typically not necessary to view the entire list, as relevant information is likely to be found on the top of the list. However, due to his general interest in international news, the user decided to get a broad overview and started to browse through the entire topic list. He continued reading the top-ranked headlines of three additional topics, covering the UK pay back discussion (second topic in Figure 5) and two other topics related to Russia and Ukraine. Having a personal interest in the Ukraine, the user started to explore coverage on the Ukraine crisis and decided to search for more information on this topic.

The user sought to get a deeper understanding of the topic by reading the articles shown in the article list (Figure 6). To further explore the topic, he opened several articles from their original websites and reviewed the complete articles.

The user reported that he got an overview about the most important topics for that day similar to the way he would have completed



	RU	UA	ALL_DOCS
GB	Ukraine says Russian military column has entered east of country	Ukraine crisis : Kiev accuses Russia of military invasion after 'tanks cross border	Russia Sends Dozens Of Tanks Into Ukraine
UA	Moscow says it 'respects' but does not 'recognize' Donbas elections	200 militants killed near Donetsk airport in past 24 hours - Ukrainian military	Ukraine expects Russia to return to practical implementation of Minsk agreements
DE	Kyiv : 32 tanks enter Ukraine from Russia	Kyiv calls Berlin amid Russian incursion reports	Germany's east still lags behind
RU	Moscow Supports Geneva Format on Ukraine Crisis, but Against Empty Meetings: Ministry	Foreign Policy Adviser Says Russia Committed to Peace Process in East Ukraine	Ukraine Crisis , Sanctions Against Russia Not on G20 Agenda in Australia : Russian
FR	Kyiv accuses Russia of sending more tanks and troops into Ukraine	German parliament kicks off weekend of Berlin Wall commemorations	Australia's G20 Presidency and the OECD : Joining forces to address the challenge of
US	Ukraine accuses Russia of sending dozens of tanks	Ukraine accuses Russia of sending dozens of tanks	Ukraine accuses Russia of sending dozens of tanks

Figure 7: Sorted matrix to facilitate topic exploration.

that task using current news aggregators. Furthermore, he was able to quickly view additional details via the stepwise expandability of NewsBird's topic and article list.

5.2 Broadening News Perspectives

After getting aware of the news situation in the first use case, the user was particularly interested in the Ukraine crisis and wanted to get a broader understanding of the invasion accusations. Using the action *Sort matrix by this topic* from the context menu in the topic list, NewsBird rearranged the cells in the matrix view so that topic-related cells were placed nearby to facilitate their comparison. Figure 7 shows an excerpt of the matrix that was displayed to the user, sorted by the turquoise Ukraine crisis topic.

The user recognized that all Western countries reported similarly on Russia's tank invasion into the Ukraine. However, the matrix also revealed to him that the invasion was much less frequently mentioned in Russia's media landscape. Instead, Russian media coverage of the Ukraine crisis primarily reported that Russia is commitment to the "Peace Process in East Ukraine" (see RU-UA in Figure 7), which is contrary to Western news coverage. He noticed that finding such information is not supported in established news aggregators. Note that the RT headline from Table 1 is not displayed in cell RU-UA, as it represents a perspective that is comparably less frequent than the ”Peaces Process” perspective that the cell in Figure 7 shows. To see details on these contradictory statements, the user hovered over corresponding cells to reveal summaries of their main articles. While each cell shows a summary of its main topic, the other topics of the cell can be viewed within the topic list described in Section 5.1.

For a deeper understanding of the differences in reporting, the user started to analyze the individual articles for corresponding cells. First, he wanted to investigate what media in Great Britain reported about the specific topic and Russia. He switched to the article list of the chosen topic, published in Great Britain and mentioning Russia. From there, he opened three articles on their original webpages to get all available information. After he realized that the major coverage in England unanimous, he switched back to the matrix visualization to see where the reporting from other countries differs. Via the context menu functionality *Open top article*, the user jumped to the most important articles covering this topic published in Germany, the U.S., and Russia. Doing so confirmed the initial finding: while Western media reported that Russian troops have already crossed the border, the headlines of Russian articles report the contrary. The user noted that Russian media claim that statements of "Allegations of Russian Troops advancing towards Ukrainian boarder" are unfounded. Even more, they insist on Russia's commitment to peace towards eastern Ukraine (see Figure 6, second article) and that the Ukrainian military does not stick to the ceasefire agreement (third article), resulting in the war forcing Ukrainian citizens to leave the country (last article).

Established news aggregators do not reveal these different perspectives as they only show media from the user's country and summarize common information. However, with the help of NewsBird, the user was able to effectively find these intra-topic differences, and thus broaden his news perspective. Ultimately, he reported that his interaction with NewsBird made him particularly aware of the effects of media bias.

5.3 Constraining the Analysis Scope

After the in-depth analysis of the different perspectives on the Ukraine crisis, the user got interested in another topic that happened in November 2014: the 25th anniversary of the Fall of the Berlin Wall. Compared to the Ukraine crisis, which was relevant to many countries, the Berlin wall topic was much less frequent in international news. Only three matrix cells contained this topic, which is not enough for the user to get a deep understanding of this topic in an international context. To broaden the understanding of infrequent topics, such as the Fall of the Berlin Wall, users can define a custom query (see Section 4.3). Via the topic list context menu *Search for this topic*, the user launched a new query using the same analysis scope as for the first two use cases, but restricted the resulting documents to the chosen topic.

In the resulting matrix, the user noted that France and Ukraine were no longer displayed as publishing countries, as they published no articles on this topic. While scanning the topics in the cells, he realized that many countries mention the former Soviet statesman Gorbachev. The user got interested in how particularly Gorbachev is involved in the 25th anniversary and changed to the article list of `RU-DE` on this topic. The first article revealed that Gorbachev opened an exhibition devoted to the Fall of the Berlin Wall.

Although the topic on the Fall of the Berlin Wall entails much less controversy in reporting than the Ukraine crisis, using NewsBird's topic control enabled the user to broaden his perspective on the topic by reading different and diverse information. Again, the concept of getting aware of perspectives in different countries is an enrichment inherent to NewsBird that also helps to identify the presence – or here absence – of (strong) media bias.

6 DISCUSSION AND OUTLOOK

While MNA is a general analysis approach focusing on the reduction of media bias, NewsBird currently only supports the exploration of international news topics. Our future research will focus on the generalization of the matrix initialization phase to support further news categories. 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 [40], or to identify text characteristics in different articles and subsequently comparing such differences on article and sub-article level, cf. [39]. A first task, however, will be to develop dimensions that enable the comparative analysis of further news categories. For instance, by dividing articles using pre-defined groups of news sources, e.g., sources that often publish the same perspective or sources that use similar slant. Such information could also be calculated automatically. Dimensions could also simply divide articles by publisher, e.g., if the user wants to get an overview of how specific publishers portray an event. This could help to reveal different perspectives for news topics that only a few articles report on, e.g., on local politics.

Correctly mapping the mentioned country is crucial for the matrix initialization phase. While a user can easily derive this information, it is nontrivial to derive via NLP. Our naive method achieves good results, but cannot handle synonyms and semantically similar phrases. We tested basic query expansion techniques using DBpedia and WordNet, but achieved mixed results. Also, we want to investigate additional use cases and visualizations for the time-range dimension.

Topic extraction is an important task of our analysis. We noticed that for matrices with $mn \gg 40$, cells with few documents will not have their own topic, but rather a topic from other cells. Increasing k helps to solve this issue, but we want to investigate different LDA configurations, such as number of cell documents and document length, too. Disadvantages of probabilistic models, such as LDA, are topic stability and reproducibility. The evaluation showed that our implementation sometimes shows imprecise, e.g., duplicate, topics. Thus, we also want to investigate other clustering techniques.

Our summarization method achieves good results, but does not fully exploit the potential of MNA. We want to 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.

MNA is built around structuring and analyzing articles and topics in a matrix, which is why the first implementation of NewsBird uses a matrix to visualize the analysis results. Our evaluation has shown that this rather simple visualization already enables users to identify and learn different information of a topic. Similar to established systems, NewsBird's matrix gives an overview of the general news situation. However, we want to examine different visualizations to improve usability. For example, for coverage on international news topics that likely differs between countries, a geo-visualization may achieve better results.

In the evaluation, we investigated the three main use cases that we see for NewsBird. The task of giving an overview about important current topics is common in news aggregation. NewsBird supports the user with a suitable topic selection and relevant details on demand. The implementation of our MNA concept supports broader news awareness by displaying different information. 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.

To perform MNA on more articles, we want to integrate news-please, an integrated crawler and extractor for news articles [21]. news-please can monitor a website's RSS feed and automatically extract the most recent articles.

Lastly, we want to investigate more use cases for MNA. We think that the MNA analysis concept can be applied to any data that expresses a relation between two entities, such as emails, e.g., the Enron Email Dataset [26], 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

In this paper, we introduce matrix-based news analysis (MNA), a novel news exploration approach that reveals different perspectives in news to reduce effects of media bias. MNA groups news articles into 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. In our evaluation, we demonstrated that NewsBird is capable of giving users an overview of news topics, while also highlighting different perspectives on those topics.

In our case study, we found evidence that the NewsBird approach broadens the reader's news perspective in everyday use cases and makes users aware of media bias and its effects on news coverage. NewsBird organizes news content with regard 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. 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 generalize NewsBird to support additional news categories.

ACKNOWLEDGEMENTS

This work has been supported by the Carl Zeiss Foundation.

REFERENCES

[1] Martin Atkinson and Erik Van der Goot. 2009. Near real time information mining in multilingual news. In *Proceedings of the 18th international conference on World wide web*. ACM, 1153–1154.

[2] Brent H Baker, Tim Graham, and Steve Kaminsky. 1994. *How to identify, expose & correct liberal media bias*. Media Research Center Alexandria, VA.

[3] Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik Van Der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. 2013. Sentiment analysis in the news. *arXiv preprint arXiv:1309.6202* (2013).

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

[5] Dan Bernhardt, Stefan Krasa, and Mattias Polborn. 2008. Political polarization and the electoral effects of media bias. *Journal of Public Economics* 92, 5 (2008), 1092–1104.

[6] Timothy J Besley and Andrea Prat. 2002. Handcuffs for the grabbing hand? Media capture and government accountability. (2002).

[7] Clive Best, Erik van der Goot, Ken Blackler, Tefilo Garcia, and David Horby. 2005. *Europe media monitor*. Technical Report. Technical Report EUR 22173 EN, European Commission.

[8] David M Blei. 2012. Probabilistic topic models. *Commun. ACM* 55, 4 (2012), 77–84.

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

[10] Business Insider. 2014. These 6 Corporations Control 90% Of The Media In America. http://www.businessinsider.com/these-6-corporations-control-90-of-the-media-in-america-2012-6. (June 2014). Accessed 20-January-2017.

[11] Erik Cambria and Bruce White. 2014. Jumping NLP curves: a review of natural language processing research [review article]. *Computational Intelligence Magazine, IEEE* 9, 2 (2014), 48–57.

[12] CNBC. 2014. Tank column crosses from Russia into Ukraine: Kiev military. http://www.cnbc.com/id/102155038. (2014). Accessed 21-January-2017.

[13] David Domke, Mark D Watts, Dhavan V Shah, and David P Fan. 1999. The politics of conservative elites and the "liberal media" argument. *Journal of Communication* 49, 4 (1999), 35–58.

[14] Frank Esser, Carsten Reinemann, and David Fan. 2001. Spin Doctors in the United States, Great Britain, and Germany Metacommunication about Media Manipulation. *The Harvard International Journal of Press/Politics* 6, 1 (2001), 16–45.

[15] David Kirk Evans, Judith L Klavans, and Kathleen R McKeown. 2004. Columbia newsblaster: multilingual news summarization on the Web. In *Demonstration Papers at HLT-NAACL 2004*. ACL, 1–4.

[16] Ilias Flaounas, Marco Turchi, Omar Ali, Nick Fyson, Tijl De Bie, Nick Mosdell, Justin Lewis, and Nello Cristianini. 2010. The structure of the EU mediasphere. *PloS one* 5, 12 (2010), e14243.

[17] GALLUP. 2015. Americans' Trust in Media Remains at Historical Low. http://www.gallup.com/poll/185927/americans-trust-media-remains-historical-low.aspx. (September 2015). Accessed 28-January-2017.

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

[19] Matthew Gentzkow and Jesse Shapiro. 2005. *Media bias and reputation*. Technical Report. National Bureau of Economic Research.

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

[21] Felix Hamborg, Norman Meuschke, Corinna Breitinger, and Bela Gipp. 2017. news-please: A Generic News Crawler and Extractor. In *Proceedings of the 15th International Symposium of Information Science*.

[22] Tony Harcup and Deirdre O'neill. 2001. What is news? Galtung and Ruge revisited. *Journalism studies* 2, 2 (2001), 261–280.

[23] Erik Hatcher and Otis Gospodnetic. 2004. Lucene in action. (2004).

[24] Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 168–177.

[25] Xiaojiang Huang, Xiaojun Wan, and Jianguo Xiao. 2011. Comparative news summarization using linear programming. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. Association for Computational Linguistics, 648–653.

[26] Bryan Klimt and Yiming Yang. 2004. The enron corpus: A new dataset for email classification research. In *Machine learning: ECML 2004*. Springer, 217–226.

[27] Christian Kohlschütter, Peter Fankhauser, and Wolfgang Nejdl. 2010. Boilerplate detection using shallow text features. In *Proc. of the third ACM int. conf. on Web search and data mining*. ACM, 441–450.

[28] Steven Kull, Clay Ramsay, and Evan Lewis. 2003. Misperceptions, the media, and the Iraq war. *Political Science Quarterly* 118, 4 (2003), 569–598.

[29] Chin-Yew Lin and Eduard Hovy. 2002. From single to multi-document summarization: A prototype system and its evaluation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 457–464.

[30] Kathleen R McKeown, Regina Barzilay, David Evans, Vasileios Hatzivassiloglou, Judith L Klavans, Ani Nenkova, Carl Sable, Barry Schiffman, and Sergey Sigelman. 2002. Tracking and summarizing news on a daily basis with Columbia's Newsblaster. In *Proc. of the second int. conf. on Human Language Tech. Research*. 280–285.

[31] M Mark Miller. 1997. Frame mapping and analysis of news coverage of contentious issues. *Social Science Computer Review* 15, 4 (1997), 367–378.

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

[33] Sendhil Mullainathan and Andrei Shleifer. 2005. The market for news. *American Economic Review* (2005), 1031–1053.

[34] Sean A Munson, Stephanie Y Lee, and Paul Resnick. 2013. Encouraging Reading of Diverse Political Viewpoints with a Browser Widget.. In *ICWSM*.

[35] Sean A Munson, Daniel Xiaodan Zhou, and Paul Resnick. 2009. Sidelines: An Algorithm for Increasing Diversity in News and Opinion Aggregators.. In *ICWSM*.

[36] Nic Newman, David AL Levy, and Rasmus Kleis Nielsen. 2015. Reuters Institute Digital News Report 2015. *Available at SSRN 2619576* (2015).

[37] Daniela Oelke, Benno Geißelmann, and Daniel A Keim. 2012. Visual Analysis of Explicit Opinion and News Bias in German Soccer Articles. (2012).

[38] Georgios Paliouras, Alexandros Mouzakidis, Vassileios Moustakas, and Christos Skourlas. 2008. PNS: A personalized news aggregator on the web. In *Intelligent interactive systems in knowledge-based environments*. Springer, 175–197.

[39] Souneil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. 2009. NewsCube: delivering multiple aspects of news to mitigate media bias. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 443–452.

[40] Souneil Park, Minsam Ko, Jungwoo Kim, H Choi, and Junehwa Song. 2011. NewsCube 2.0: An Exploratory Design of a Social News Website for Media Bias Mitigation. In *Workshop on Social Recommender Systems*.

[41] Souneil Park, Minsam Ko, Jungwoo Kim, Ying Liu, and Junehwa Song. 2011. The politics of comments: predicting political orientation of news stories with commenters' sentiment patterns. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 113–122.

[42] Dragomir R Radev, Hongyan Jing, and Malgorzata Budzikowska. 2000. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. In *Proceedings of the 2000 NAACL-ANLP Workshop on Automatic summarization*. Association for Computational Linguistics, 21–30.

[43] Dragomir R Radev, Hongyan Jing, Ma lgorzata Styś, and Daniel Tam. 2004. Centroid-based summarization of multiple documents. *Information Processing & Management* 40, 6 (2004), 919–938.

[44] RT. 2014. Moscow to Kiev: Stick to Minsk ceasefire, stop making false 'invasion' claims. http://rt.com/news/203203-ukraine-russia-troops-border/. (November 2014). Accessed 25-January-2017.

[45] Gerard Salton, Anita Wong, and Chung-Shu Yang. 1975. A vector space model for automatic indexing. *Commun. ACM* 18, 11 (1975), 613–620.

[46] Alison Smith, Timothy Hawes, and Meredith Myers. 2014. Hiérarchie: interactive visualization for hierarchical topic models. *Sponsor: Idibon* 71 (2014).

[47] Joseph D Straubhaar. 2000. *Media Now: Communication Media in Information Age*. Thomson Learning.

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

[49] University of Michigan. 2014. News Bias Explored - The art of reading the news. http://umich.edu/newsbias/. (2014). Accessed 27-January-2017.

[50] Wayne Wanta, Guy Golan, and Cheolhan Lee. 2004. Agenda setting and international news: Media influence on public perceptions of foreign nations. *Journalism & Mass Communication Quarterly* 81, 2 (2004), 364–377.

[51] Marcos Weskamp. 2016. newsmap. http://newsmap.jp/. (2016). Accessed 10-January-2017.