

Review

Not peer-reviewed version

---

# Mining Society's Snapshot: A Taxonomy and Survey of News Analytics

---

[Olusola T. Babalola](#)\*, [Bolanle Ojokoh](#), Olutayo Boyinbode

Posted Date: 5 February 2026

doi: 10.20944/preprints202602.0243.v1

Keywords: news analytics; text mining; taxonomy; computing for social good; information extraction; survey



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

# Mining Society's Snapshot: A Taxonomy and Survey of News Analytics

Olusola T. Babalola <sup>1,\*</sup>, Bolanle Ojokoh <sup>2</sup> and Olutayo Boyinbode <sup>3</sup>

<sup>1</sup> Department of Mathematics and Computer Science, Elizade University, Ilara-Mokin, Nigeria

<sup>2</sup> Information Science Department, Federal University of Technology, Akure, Nigeria

<sup>3</sup> Information Technology Department, Federal University of Technology, Akure, Nigeria

\* Correspondence: olusola.babalola@elizadeuniversity.edu.ng

## Abstract

News analytics has traditionally been dominated by financial market prediction; however, its application across broader societal domains is vast and fragmented. This paper presents a comprehensive survey and a novel taxonomy of non-market news analytics research. By analyzing 309 research entries through an iterative open-coding process and aligning them with DDC and IPTC standards, we identify 25 distinct research categories including public health, civil unrest, and environmental sustainability. This work provides a structural baseline for understanding how automated news analysis serves as a computational tool for social good.

**Keywords:** news analytics; text mining; taxonomy; computing for social good; information extraction; survey

## 1. Introduction

News can be considered a snapshot of a representative portion of a society at a certain period of time. Historical and current news therefore has been of interest to researchers across wide-ranging fields. Some understudy how the society, or a segment of it is depicted in the media, others are interested in the perception of the society about one thing or the other, and some others are interested in deriving valuable insights about the society as reported in a news collection. Notable research areas with interests similar to the ones mentioned are content analysis[1,2] and news analytics[3] research. This survey targets news analytics.

We highlight a distinction in news analytics research. Our observation, which serves as motivation for this work, is that a distinct and popular stream of research is present in news analytics - that of financial news analytics. We refer to that stream as "news analytics for the markets". The primary concern of financial news analytics is predicting prices or price movements of financial products based on the tone of news reports about companies, their performance and situations among other considerations. The prediction is not limited to company stocks only but also to commodities, oil futures etc. The interest, visibility and high count of research works directed at the markets has resulted in bibliographies, surveys, and books devoted to the topic.

Despite the picture painted by the visibility and abundance of financial news analytics works, several other news analytics applications exist, as a substantial body of literature shows, with those applications targeting diverse spheres of the society. The question which other news analytics research domains exist?' prompted this work.

## 2. Methodology

In this work, we consider news of the electronic type, published by a recognized newspaper with editorial guidelines, rather than citizen-published such as on blogs and social media. We included works based on data derived from news articles e.g. titles, and those utilizing comments about the contents of the news article; in short, the research work must ingest news or a part of the news as data.

We categorize works by considering the motivation, purpose or objective of the work, as well as the results and outputs of the analytics system. If there is a keyword list in the article, we consider the keywords to determine the overall theme and the category of the work. While some themes such as health can be identified in a straightforward manner, some others are not easily labeled. The entire work determines whether or not a work is added to a theme, and not the specific approaches and techniques used (methodology). The center of the entire work must be news-related.

As expressed by the title of our article, we exclude all research aimed at the markets stock markets and financial markets. Excluding market-focused work does not preclude business and economics from featuring on our taxonomy as there are other applications of news analytics in business aside from predicting the markets. A category is formed if at least one news analytics work is found in a different domain than previous works. The domain must be a major human activity or interest otherwise the work is categorized as "Others". An article may belong to more than one category. Where this is the case, we simply add an additional entry for such in the secondary theme.

### 2.1. Literature Search

The primary tool used for discovering the works examined was Google Scholar (<https://scholar.google.com/>). Searches were majorly conducted between December 2018 and February 2020. Articles authored in English language were considered and examined.

The set of queries to the academic search engine, which formed the bulk of the literature search are listed in Table 1. For each of these queries, the first 500 results were examined. The queries in Table 2 form the secondary set of queries. For those queries, we examined at least the first 100 results. In a limited number of instances, we explored the references in discovered articles to find other relevant works.

We chose Google Scholar as it is one of the largest, open, academic search engines for scholarly work discovery. The assumption is that a lot of researchers carrying out literature search are likely going to include Google Scholar or the Google search engine in their search strategy due to the fact that no subscription or registration is needed before Google Scholar could be used and because it is integrated with the most popular Internet search system. Could this decision to use a single tool limit the works discovered? Maybe. However, previous empirical studies showed a good coverage by Google Scholar while comparing it with other scholarly databases [4,5].

The literature search did not discriminate against any form of scholarly work since our objective was to discover the breadth of news analytics in academic research. We considered thesis and dissertations as long as they matched the inclusion criteria.

**Table 1.** Search queries used as the major query in the literature search

Main Search Queries
"News analytics" -finance
"Text analytics"
"Analyzing news" or "Analyzing news"
Text mining and news

**Table 2.** Search queries used as the supporting query in the literature search process

Supporting Search Queries
News analytics
News analysis
Analyzing news

#### 2.1.1. Search Method

From the over 3000 results obtained in the literature search, we first make a pre-selection based on the titles and search result snippets. This selection is to eliminate those articles where the titles and snippets clearly show they do not meet with the criteria for inclusion. We proceeded to carry out a critical examination of the abstracts of 780 articles selected from whittling the results from the

literature search. Articles whose abstracts show a discernible purpose matching the inclusion criteria are selected. For other articles, we examined the methodology and the full text of the paper to have a clearer understanding of whether the article satisfies our inclusion criteria. At the end of this process, we gathered articles that could be categorized as non-market news analytics research; the result is 309 entries in total over 25 categories.

### 2.1.2. Purpose of this Research

Our aim is to explore the knowledge domains and aspects of the society that news analytics has covered, present representative works to increase the visibility of the non-markets news analytics segment, and attempt to show the frontiers of news analytics applications that are exposed within the scope of our literature search.

### 2.1.3. Criteria for inclusion

Certain basic conditions were used to determine inclusion of a research work in this survey.

- Firstly, some algorithmic processes of computation must be involved in the methodology of the research work.
- Does the research solve a news analytics problem as an end or is a means to an end? The news analytics must not be part of the "method" or a component in the overall work. Rather the news analytics should be the aim or objective of the overall work.
- The processes must manipulate a news dataset or a news-derived dataset.
- The work should include key processes that are common in text mining such as those mentioned in [6,7], and those used in news analytics tasks[3]. Therefore, we are concerned with analytics via an automated process executed by a non-human computing agent.

### 2.1.4. Literature Search Limitations

Despite ensuring sufficient depth when using Google Scholar, our literature search is not in any way considered as exhaustive or conclusive.

## 2.2. Categorization

Determining categories for each considered article is straightforward in some cases, and in substantial cases it is challenging.

We opt for utilizing as many categories as possible and allow placing an article into more than one category. This safeguards against forcing an article into a category. It also serves for articles reporting research that are clearly intended or designed to achieve more than one objective.

Rubrics used for determining which category to place a work is as follows.

- What knowledge domain is the result of the research in? If the output of the research is predictive, the predicted information may be classified into one domain (category) or the other.
- Is the dataset general or domain-specific? In some cases, the dataset used is from news of a specific category. This observation is helpful in categorizing such research.
- Are there keywords in the article? In most of the papers we discovered an absence of a keyword list. The titles, and abstracts of the articles however could mention keywords that define which category a research fits.
- It may be possible for a research to fit more than a category. For example, an exploratory information system for health-related events would simply be grouped in the health and information presentation categories.

### 2.2.1. Coding Guidelines

We adopted open coding [8], one author assigned articles to categories until no new categories were needed. During the coding, new categories were added and when necessary merged into larger categories. Iterating qualitative coding, by adding and collapsing categories offers improvement in the quality of the categories [9].

In addition to the open coding technique, we use category labels drawn from the vocabulary of two notable classification systems from different domains (Section 2.2.2). There are over 3000 categories and subcategories in all from the DDC and IPTC combined, this grants a perfect range of labels to choose from.

### 2.2.2. Naming the Categories: Library Categorization List + News Categorization List

In order to assign names to categories, we utilize knowledge classification systems used in library organization combining it with classification systems for news organization. Utilizing both is because the scope and objective of news research varies, ranging between knowledge in academic and social domains; we are therefore able to find headings for our categories from either of the classification systems. Our category labels are obtained from a mix of the DDC and IPTC vocabulary.

The Dewey Decimal Classification (DDC) [10] was developed to be ideal as a general knowledge organization tool and is used in library organization. The IPTC subject codes was designed to standardize all news content classification. It is focused on text and consists of 1,400 terms structured into three levels.

When the coder decides on a category, she consults the International Press Telecommunications Council (IPTC) subject codes first, if there is a category or subcategory with that label, then that category becomes the label. If it does not exist in the IPTC, the Dewey Decimal Classification (DDC) is checked. The IPTC subject category is standardized for classifying news articles. Certain articles are not social-oriented and so may not have a matching IPTC category entry. In that case, the DDC is a general knowledge organization tool that provides an alternative.

**Table 3.** Count of works by category

SN	Category	Publications
1.	Crime	18
2.	Unrest/protest	25
3.	Politics	24
4.	Healthcare	24
5.	Business/Economy (excluding the markets)	24
6.	Future prediction/Forecasting (e.g. predicting future events)	8
7.	Historical happenings and history	3
8.	Public perception	3
9.	Armed conflicts/violence	6
10.	Terrorism	8
11.	Disasters/Risks	13
12.	Environment/Sustainability	10
13.	Information exploration and Visualization	34
14.	Technology	2
15.	Journalistic, linguistic, analysis of news senses. The interests include the comparison of news coverage across news outlets or comparison of news coverage across countries, analysis of bias etc. (The research target in this category is to examine beyond the news topic itself to "making" or production of the news)	20
16.	Notoriety e.g. hate speech	2
17.	Gender issues	2
18.	Gambling	1
19.	Migration	7
20.	Sports	1
21.	Societal/government. This category examines issues across different societies including government actions	16
22.	Emerging and interesting entities. This is not Named Entity Recognition (NER) but rather the identification or discovery of important and notable events in the news	22
23.	Negotiation	1
24.	Education, science and technology	3
25.	Miscellaneous applications	21

### 2.3. Categories

We proceed to present the categories discovered from the selected news analytics works. We explain those category headings that may not be clear or that are ambiguous using one of the articles in that category. The categories are presented in the following list; the position of a category in the list is insignificant.

1. Crime
2. Unrest/protest
3. Politics
4. Healthcare
5. Business/Economy (excluding the markets)
6. Future prediction/Forecasting (e.g. predicting future events)
7. Historical happenings and history
8. Public perception
9. Armed conflicts/violence
10. Terrorism
11. Disasters/Risks
12. Environment/Sustainability
13. Information exploration and Visualization
14. Technology
15. Journalistic, linguistic, analysis of news senses. The interests include the comparison of news coverage across news outlets or comparison of news coverage across countries, analysis of bias etc. (The research target in this category is to examine beyond the news topic itself to "making" or production of the news).
16. Notoriety e.g. hate speech
17. Gender issues
18. Gambling
19. Migration
20. Sports
21. Societal/government. This category examines issues across different societies including government actions.
22. Emerging and interesting entities. This is not Named Entity Recognition (NER) but rather the identification or discovery of important and notable events in the news.
23. Negotiation
24. Education, science and technology
25. Miscellaneous applications

## 3. News Analytics and Text Analytics: The Applications, Techniques, and State of the Art in Brief

"Analytics is the discipline of applying mathematical sciences to data for the purpose of making better decisions. Analytics serves to convert the increasing amounts of data into actionable information" [11]. Analytics is the process of analysis of data that is done logically and aided by sciences (e.g. mathematical and statistical sciences, Computer Science). Analytics can be descriptive, diagnostic, predictive or prescriptive[12].

The analysis of text content in emails, blogs, tweets and other forms of textual communication constitutes what we call text analytics. Text analytics has also been called text mining. Text analytics is an extension of data mining that tries to find textual patterns from large non-structured sources as opposed to data stored in relational databases.

Analytics involves the use of computational approaches, mathematics, and statistics to process data in order to extract fine information or insights which can form the basis of decisions and actions. The analytics process often works on a large collection of data, for example a corpus of news articles

comprising news from several decades, and produces insight that is easily grasped in return. The application of analytics on news content data can be regarded as news analytics.

Das opined multiple levels at which news analytics could occur; "We may think of news analytics at three levels: text, content, and context. Text-based applications exploit the visceral components of news i.e. words, phrases, document titles, etc. The main role of analytics is to convert text into information. This is done by signing text, classifying it, or summarizing it. A third layer of news analytics is based on context. Context refers to relationship between information items. Google Pagerank algorithm is a classic example of analytic that functions at all three levels. The algorithm has many features, some of which relate directly to text. Other parts of the algorithm relate to content, and the kernel of the algorithm is based on context, i.e, the importance of a page in a search set depends on how many other highly ranked pages point to it. Indeed search is the most widely used news analytics"[3] .

They went further to define news analytics. "The term 'news analytics' covers the set of formulas, and statistics that are used to summarize and classify public sources of information. News analytics is a broad field, encompassing and related to information retrieval, machine learning, statistical learning theory, network theory, and collaborative filtering"[3] .

Text mining is a set of procedures and operations applied on textual content which is unstructured in nature, with the objective of determining the presence or absence of patterns that may assist in arriving at a desired conclusion or provide new information. Text mining is applied on written text in electronic form such as emails, blog posts, web pages, social media posts, tweets, company reports and documents, textbooks, and scientific publications. However, text mining is not limited to textual communication alone; content having the same format as written text such as DNA data has also been analyzed using text mining techniques.

Some authors are of the opinion that text mining is the same as text analytics thereby using the two interchangeably, while some have a different view. Authors, including Kent say there is a distinction. In their opinion, text mining considers text as words and extracts the count of words in a document, the type of those words, and the number of kinds of words within kinds of documents (the document collection). Analytics on the other hand extracts quality information from unstructured data in order to distill meaning. It goes beyond counting words to extracting meaning and offering some context to that meaning[13] .

### 3.0.1. News Analytics System Framework

We use the diagram in Figure 1 to give a concise description of a news analytics system.

- A1 is a single document source, and A2 a multi-document source.
- B is a set of operations that manipulate or process electronic text from A1 or A2 on an application (use-case) basis.
- C is a set of data derived from A1 or A2, and in a form that can be processed or analyzed by subsequent computational processes.
- D is a set of algorithms or operations designed to mainly mine data or obtain insights.
- E is the result of the analytics process; information, or insights.
- F is a visualization process.
- G is a process which ingests the results of the analytics.
- $\phi$  is external data which may be utilized by the analytics system as a domain knowledge source.

B is the commonly referred to "preprocessing step" in text mining systems. D is the major analytics stage, and the specific algorithms or processes found at this stage are application- or domain-dependent. A good coverage of what common problems are tackled at this stage, the algorithms which are commonly used, and details of news analytics processes in general can be found in [6,14,15].

Text mining, with some data mining, is practically what news analytics systems apply; therefore any good resource on text mining would likely contain details of what operations may be carried out in the stages identified in Figure 1.

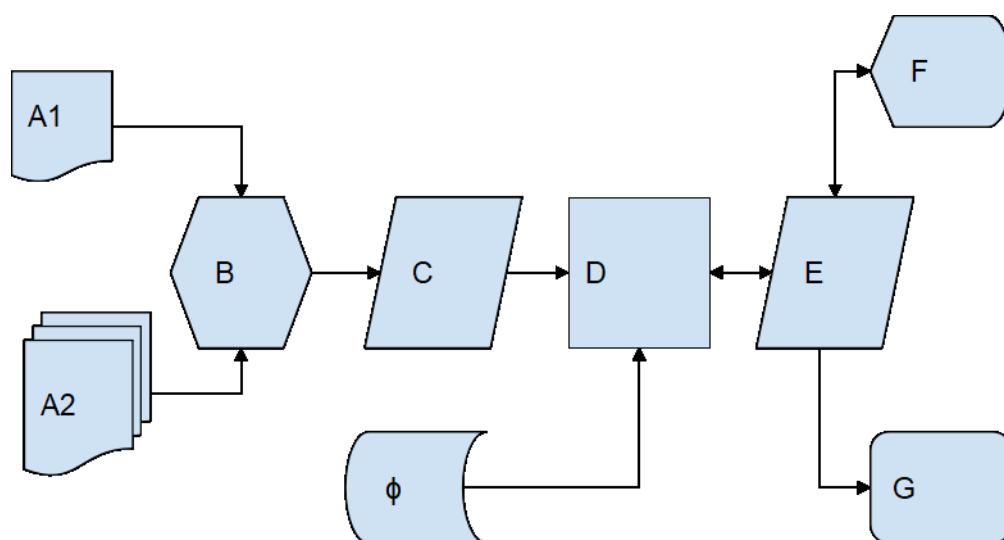


Figure 1. Architecture (flowchart) of a typical news analytics system

Another major stage that can be found in a certain category of news analytics systems is the visualization stage (Stage E). In some applications, the visualization stage (Stage E) may replace the main analytics process stage (Stage D), thereby becoming the main analytics process that is present in the system. Section 4.13 is devoted to news analytics works that are visualization-based.

#### 4. Non-Markets News Analytics

We proceed to present non-markets news analytics in detail. The works presented in each of the following categories are arranged in chronological order from the oldest to the newest. The arrangement is only according to the year of publication and not fine-grained to the month or day of publishing; therefore within our presentation, it may be possible to find a work published earlier in a year showing after a later publication. Where there are multiple literature devoted to the same work with some arising at a later time, we simply cite them as a set with the first mention of the work in the category they belong to.

##### 4.1. Civil Unrest, Protests, and Strikes

News mining might have predicted a political uprising such as the Arab Spring, as was found out by a retrospective study of the 'tone' of the news in Egypt, Libya, and Tunisia over several years[16]. The utilization of news mining to infer political rifts is not uncommon, rather it is a key goal of computational social science and culturomics[17]. The predictive nature of social media news in highlighting the causes and evolution of civil unrest was discussed in [18]. From analyzing the GDELT dataset [17] reported that political instabilities with global impact such as the Arab Spring are associated with breakdown or enhancement of long-range correlations in political events. Their work provides a way for researchers to observe important news events. In the case of political crisis at a national level, [19] mined political crisis signature from GDELT data and used it to forecast political crises. A different approach to forecasting protests was proposed in [20] which was based on dynamic content search followed by extraction of future temporal references from content found from news and social media data. This system was tested for 10 countries in Latin America and was able to capture unrest with a lead time of about four days.

The EMBERS system[21] operates on a similar scope as [20]; EMBERS is a major project born from an industry-university partnership. EMBERS undertakes continuous monitoring of media and forecasting of social unrest across Latin America. It is also capable of informing on other events aside from civil unrest events. Data from a range of sources such as Twitter, news and blog feeds, and alerts from Healthmap are some of the many data sources EMBERS mine. The project utilizes various methodologies in the acquisition of data, processing of the data, mining and presentation of insight



from its processes including diverse functionalities can be found in the series of literature on this project[22–33].

Discriminating patterns for detecting and extracting occupy movement protests from GDELT data were examined in[34]. The PALOMAR system [35], a computational journalism tool, mines a collection of public and private datasets to allow journalists to track interesting events including protest events. Automated detection and extraction of security-related events including political violence and social unrest was considered in [36] for the purpose of creating a corpus of such events. The Carbon system[37] which was built to forecast civil unrest for Australia and a number of Asia-Pacific countries addressed some shortcomings of EMBERS as regards coverage, access or licensing, and fine-grained prediction. A similar system is reported in [38]. In Southeast Asia, the GDELT was mined to understand the development of social unrest; the system has been experimented on data from five countries in the region and shows effectiveness.

In a different motivation, the GDELT data was analyzed by [39] towards understanding the actor dynamics in ethnic conflicts and human right violations. Another interesting civil unrest related research question in the news mining literature is the determination of what turns mass gatherings violent [40], in that research, data from various news sources was used in forecasting violent behavior in crowds. The system for extracting protests from news in [41] was developed to predict what news led to a protest. Statistics: 25 works.

#### 4.2. *Crime, Legal*

In compiling their electronic repository of expert witnesses [42] extracted information about candidate professionals from newspaper articles automatically. Cross-border crimes including human-trafficking, smuggling, illegal immigration and crisis events are automatically extracted from online news in [43]; a similar research described as the Real-time News Event Extraction Framework [44], is devoted to extracting structured information on security issues, including smuggling and violent events from online news. Extracting and clustering crimes by type from Arabic newspapers is described in [45]. A more robust work designed to serve as a news monitoring system in Malaysia for the crime domain [46] tracks, analyzes, organizes, and presents insights from crime stories. The gender profile and age-ranges of offenders was obtained from crime news in [47]; a system which was implemented for large scale quantitative narrative analysis (QNA).

An approach that can be applied in mining large collections of crime news to aid crime exploration and investigation was demonstrated in [48]. Similarly, the detection of crime patterns in news was discussed in [49]. A crime profiling system in the Arabic language domain was proposed in [50], the system mines type of crime, where the crime occurred and the nationality of the actor(s) from crime news reports. Like other reported works, data from public sources are exploited to create patterns and trends of crimes by region in [51]. Among a host of other features, [52] presented crime hotspots mapping with visualization based on newspaper articles; another crime hotspot related research [53] predicted where the crime hotspot of the future year will be. This work, [53] did not use newspaper articles as its datasource but used a crime dataset instead. However integrating text mining into the methodology described is straightforward and doing so will increase the versatility of the system. Our survey of news analytics in the crime domain did not reveal any research work that considered further utilization of insights drawn from the analytics operation, and the task of predicting the crime outlook of the future is certainly an interesting challenge. Airport threats were mined from news and social media sources in [54] with the objective of assessing and ranking potential threats each airport face.

Text mining may deepen the scope of crime analysis; since more than 50% of crimes are unreported to the police, analyzing sources of crime data outside police databases is desirable. The previous assertion was the motivation in [55] whose focus was analyzing newspaper reports to map crime distribution across Tanzania. To understand identity theft, [56] analyzed online news stories and reports related to identity theft, with the intent of examining behavioral patterns and resources used in these crimes. The extraction and curation of crimes into a crime register from news articles is

described in [57]. In [58] theft stories in newspapers are used in deriving various insights including crime mapping information. Statistics: 18 works.

#### 4.3. Politics

News media has been used to forecast the response of the public to political candidates [59]; how news impacts public perception of candidates was analyzed using computational linguistics techniques (part of the family of technologies used in text mining). Volume and velocity, both key nature of big data, manifests in news media; the challenge of volume is daunting in the case of news because it is primarily produced for human consumption and often contains important information that is sometimes beneath the surface. Therefore, to bridge the gap of easy consumption of political news around the 2008 US presidential election period, a visualization and visual analysis system was developed by [60], assisting political analysts in drawing conclusions without reading through every word of several news articles. In another vein, the political network of Presidents Clinton and G.W Bush comprising the social network of their cabinet members was automatically extracted from news articles in [61]. In Indonesia, the non-government organizations and institutions that are counter-radical, their characteristics, leadership, practices and sentiments were extracted from news articles [62]. To predict the political orientation of news articles, [63] analyzed commenters' sentiment patterns towards these articles. The idea behind their approach is to bypass the complex text analysis that would have been required if the actual news content were analyzed.

The DARPA Integrated Crisis Early Warning System (ICEWS) project, a large-scale automated coding system for automated production of political event data from news sources is described in [64]. At the time the research was reported, over 75 regional and international news sources are processed in near-real-time by the system. An example of mining news to measure a leader's public space approval in real-time and in response to current situation e.g. scandals or economic crisis, from the media is reported in [65]. The potential of news to predict political power challenges and shifts such as the Arab Spring was discussed in [16]. [66] showed the utility of text mining to discourse analysis using the 2007 Kenyan elections and the post-election crisis as focus; they compared local and international (UK) news reports on the election. To understand polarization in the US political space, political blogs and their comment sections were analyzed by [67]. They predicted the response of different political groupings to the same news story.

The involvement of foreign nations and their politicians in key political incidents in Egypt, Libya and Sudan over a time period was analyzed from news sources by [68]; important topics, persons, organizations, and locations in meta networks mined from the news were identified. Similarly, [69] a news analytics system was built to analyze news coverage during the 541 days of negotiations leading to the formation of a government in Belgium in 2011. It considered how the tone of reporting evolved on a party, politician and newspaper level. In [17], political events data from the GDELT were mined to understand the making of political instability globally.

Much of politics is expressed in words but the volume of political text is a challenge to scholars using texts to make inferences about politics [70]; automated text analysis is promising in this aspect. In contribution to the application of analytics to political science [70] surveyed a wide range of methods including techniques to validate the output of models.

The analysis of news to determine the feelings of the public to a political figure is discussed in [71]. Dissatisfaction with rulers and governments leads to domestic political crises (DPC). The use of GDELT data in examining the signature of DPCs is presented in [19]. The use of political events data in political forecasting was examined in [72]. To carry out sentiment analysis of political text in under-resourced languages [73] proposed an approach which is suited for analysis in such situations. The inter-mix of government and business objectives in news by state owned media in China was studied using text analysis in [74]. Also on China, the interactions between China and the rest of the world and the evolution of that relation was modeled from GDELT data [75]. Political bias in the media was studied in [76] by examining a dataset of two British newspapers. A political news dataset and

use cases including the discovery of news narratives from political news was presented in [77]. The prediction of political violence from newspaper text was considered in [78].

The dynamics of multinational interaction as a result of the South China Sea Arbitration was studied using events extracted from GDELT. The authors were of the opinion that studying post-event dynamic evolution of international relations through quantitative analysis is important [79]. Can one predict whether or not uncivil discussions will be generated from news articles? [80] discussed a method to predict what tone readers' comments will take by analyzing just the text of a news article. The polarization of news and social media during political campaigns was studied by [81]. Statistics: 24 works.

#### 4.4. Healthcare

Public health interests are not left out of content analysis, for example the study of framing of obesity in news [82]. But the use of analytics in the health domain is desirable just like it is in the other domains we have covered since an automated system would analyze more data and at a wider scope too. For instance, the work on obesity utilized a selected collection of 136 articles, exhibiting limitation volume-wise. The Europe Media Monitor (EMM), an EU-centric project which gathers news in over 50 different languages included an analytics component for health event detection and monitoring [83]. The EMBERS (Early Model Based Event Recognition using Surrogates) project is a similar monitoring project which ingests data from heterogeneous sources and automatically produces forecasts of significant societal events including disease outbreaks in real-time.

The application of text mining in health research includes Newsmap [84] an hierarchical knowledge map which was automatically generated from Chinese news in the health and finance sections. Tracing the evolution of a public health event automatically, [85] characterized the 2005 asbestos-linked health crisis in Japan using text mining and network analysis on news content. Syndromic surveillance is important in safeguarding against infectious disease outbreaks, and the application of automatic news monitoring and classification for the purpose of tracking disease outbreaks is the objective of [86]. The relationship between the media and the consumption of illicit drugs was studied in [87]. The authors analyzed news content to determine a quantitative relationship between media reports and opioid abuse.

A significant number of applications using the World Wide Web as signal source also exist for the purpose of health surveillance, a systematic evaluation of such systems was carried out in [88] using text mining. The use of social media to characterize terrorist activities and their relation to epidemic outbreak was the subject of [89]. Epidemic surveillance is again the aim in [90], however this time, the objective is to analyze multilingual news to characterize epidemic events. In what may be considered a mental health research, a system to automatically determine what emotions are induced in readers of a news article was considered in [91].

Knowle [92] extracts news events from the web and links the events according to semantic relations. The use of the Knowle system in organizing and mining health news is studied in [92]. To study whether disease coverage in newspapers and social media is actually disease detection or public opinion reflection, [93] compared measles-related social media messages, newspaper articles, and actual reported cases. Two interesting things would be highlighted about this work, the first involves the aim. A lot of web-based disease surveillance systems exist, therefore it will be beneficial to understand what the dynamics, range, and context of disease coverage in online media is. Such knowledge will help in developing better web-based disease surveillance systems. The second noteworthy point is that the analysis was manual; sentiment analysis was coded by two human raters, and the topic classification was also manual. We included the work based on its important aim and the fact that the main methods used, even though manually implemented, are established tasks in automatic text analytics.

SourceSeer [94,95] studied the outbreak prediction of rare diseases such as Hantavirus from analyzing news sources. The system forecasts the emergence and progression of epidemics. Another disease outbreak informatics system, DEFENDER [96] integrates data from social media and news for

epidemiological insights and forecasting. With their Dis2Vec system, [97] generated disease specific word embeddings from health-related news corpus and used those embeddings to automatically detect and characterize emerging diseases. The coverage of disease outbreaks on Twitter as compared with the coverage in the news, was tested in [98]. The authors were motivated to investigate how coverage differs between news providers (traditional news media and microblogging services) and used automated analysis to carry out these investigations. Different approaches to utilizing internet news sources in building generic disease surveillance tools that are applicable to tracking rare, endemic, and emerging diseases globally are investigated in [99].

The relationship between news coverage and infection incidences is studied in [100]. Dengue news articles are analyzed from news across Asian countries with India as key study in [101], topics in these news are determined and used in the construction of topic evolution graphs through the year. [102] undertook a multi-dimensional analysis of news media to reveal the coverage of diseases in news articles spanning over 30 years. A tool to collect and analyze food reports from around the world was presented in [103]. [104] used GDELT news event data to model the measurement of factors that enhance disease spread, the case study of this work was the Middle Eastern Respiratory Syndrome (MERS). Understanding news media coverage of women's health motivated the study by [105] which focused on Korea. The problem of forecasting drug safety was studied in [106]; the study focused on South Korea. Statistics: 24 works.

#### 4.5. Business and Economy (Excluding Markets)

News analytics for business ends are also found in literature. [107] mined Chinese industrial news to extract companies, events, temporal, location, and person information. A system named Knowledge-Based News Miner (KBNMiner) analyzed news information to automatically predict interest rates [108]. The extraction of events from news for the purpose of enhancing the knowledge about an organization's operating environment (environmental scanning) was researched in [109]. The automatic generation of an hierarchical knowledge graph from Chinese news in the health and finance categories is described in [84]. Another application of news analysis is found in the automatic extraction of information from news and other Internet sources to feed a negotiation agent with the contextual information needed to function effectively [110]. Analyzing news makes it possible to infer the social importance of events; in [111] the occurrence of such events as severe weather conditions have been leveraged to predict energy demand or pricing.

The automatic analysis of business-related information reported in news such as contracts won and product launches could be of value to an organization in the provision of useful insights [112]. However, this can only be achieved with the availability of rich semantic representations that are good enough for an inference engine to reason over. Capturing rich semantic representations from news is therefore the subject of [112]. Analyzing trends in the operating environment of a business aids managers to respond and adjust strategies, mining changes caused by events from news stories is discussed in [113]. [114] explored methodologies to extract competitive intelligence from news and other online sources using text mining, the authors showed how this might be further integrated with structured data for decision making. A variety of events impact on oil demand and prices; [115] detects such events using text mining of news articles, and also to determine which event might impact oil markets. Summarizing Turkish economic news and structuring it into quantitative representations [116] is partly motivated by the fact that there is too much economic news to read; news analytics systems such as this can be of importance to business managers and analysts in providing concise insights. By gathering news and other data from the web and analyzing key events found in these news, exchange rates are predicted in [117]; this work has been used in real world situations.

In other applications, the problem of automatically connecting mentions of brands to relevant business news about them, was tackled in [118]. Motivated by the fact that business events differ from other events, [119] proposed a method for detecting and categorising business-critical events in news. Environmental happenings around a business may include risks - events that could hamper a company's growth; a framework to process textual data and news documents automatically to extract

and classify risks is demonstrated in [120]. The Chinese government's usage of state-controlled media is studied by analyzing news articles in [74]. Narrative visualization is a technique to allow readers to determine what information they consume. Automatic narrative visualization and the analysis of financial news to show the socio-economic relationship between price and other topics is covered in [121]. A model developed in [122] was applied to study financial risk from analysis of news articles, with the authors showing how text can be a useful complementary source for risk analytics.

Supply chain relations are crucial to an organization but these relations are unapparent [123]; as a result, the integration of text mining and social network analysis to analyze news is considered and reported as an approach to solving the problem. The discovery of business industry structure and business relations from news with both macro- and micro- viewpoints is studied in [124]. The extraction of structured representations of economic events from news articles was undertaken by [125]. Big data analysis of the gambling industry in South Korea was undertaken by [126] to understand the press' view and the public's perception of gambling-related news. Sustainable supply chain management in the textile industry was studied in [127] by analysing news articles and sustainability reports of firms. The usage of sentiment analysis, a key technique in text analytics, in the analysis of the tourism industry and the unique domain-caused challenges such an approach face is studied in [128]. The study examined the design of sentiment analysis based works in that domain. News has also been analyzed to define indicators for bank risk culture [129]; the indicators were extracted using a variety of text mining techniques. Statistics: 24 works.

#### 4.6. Predicting the future

News articles contain information of some events that are planned to occur in future. This information is important in understanding how specific events might unfold. The task of retrieving and ranking these information from news articles was defined and addressed by [130]. Future-related information mining aids decision-making [131]; mining news for future events defined both spatially (place) and temporally (time) was described in [131]. The work by [132] takes a different path to future-prediction than the previously mentioned ones - the 'Pundit algorithm', which generates a plausible future event that might occur given a present news event. This algorithm involves mining over 150 years of news. To determine what the events outlook for the future might be, [133] researched the automated extraction and generalization of event sequences from news; learning to forecast forthcoming events from corpus containing 22 years of news articles.

A visual approach to the mining of news to provide insights into the future is described in [134], the research considered contextual similarity in their analysis as against spatio-temporal analysis only. According to [135] "text mining provides a solid base for reflecting on possible futures". The authors therefore explored the potentials of text mining, over a variety of data sources including news, for the purpose of foresight. [136] considered a news prediction technology with some focus on efficiency. Although this section is concerned about predicting events, we conclude with a unique prediction-related problem, that of cross-checking whether predicted events in news occurred. In [137] the authors built an automated predictions validation system to cross-check predictions made by journalists in prior news articles. Statistics: 8 works.

#### 4.7. History

How the past was remembered is the thrust of [138], the authors applied computational and statistical tools to analyze news covering different countries, showing how interesting topics and hidden correlations surface with the use of such tools. Past content constitutes a rich research repository; [139] investigated a system to semantically search news articles, with extended capabilities such as the identification of contrasting and contradicting information in news articles. The content in news from decent sources is relatively correct due to painstaking efforts made by journalists and such contains facts about key happenings, [140] investigated tools and methods to extract geospatial data from unstructured text archives. They described the construction of a database of historical flood events and the extraction and usage of the rich geospatial data in news sources. Statistics: 3 works.

#### 4.8. Public Perception

The influence of political news articles on the opinion of people (commenters) was examined in [63] where the authors predicted the political orientation of articles by analyzing comments. The creation of a geospatial visual analytics system which transforms text into emotional heatmaps with spatial awareness is described in [141]. The coverage of various diseases in the media and the tone (sentiments) of the disease-reporting news article was analyzed in [102] for each disease. The objective was to capture public opinions towards these diseases and to inform unmet medical needs of the populace. Statistics: 3 works.

#### 4.9. Armed Conflicts/Violence

This category of news analytics is similar to the civil unrest and terrorism categories; keeping these categories separate is because of the existence of some works that are distinctly focused. NEXUS [142], a project of the Joint Research Center of the European Commission is an event extraction system for populating knowledge bases with records of violent incidences by extracting security-related facts from news articles. For forecasting political violence at sub-state levels in Afghanistan, [143] utilized GDELT data in conjunction with spatial analysis. For a global scale analysis of armed conflicts and major protests globally between 1979-2014 [144] analyzed GDELT data across different scales and timelines. The use of analytics for the monitoring of arms and weapons including biosurveillance is briefly surveyed in [145]. The network of actors in ethnic violence was studied by analyzing data from GDELT in [39], and in [78] a methodology to predict armed conflict was presented. Statistics: 5 analytics works (and 1 review).

#### 4.10. Terrorism

The interest in [146] is the understanding of how the major events of a terror attack evolved, the authors examined the problem by drawing on CNN news articles and analyzing them. [147] worked on a similar problem with their method for tracking terror and violence in real-time from news streams. A study of counter-radical organizations, their characteristics, and networks via news analytics is presented in [62]. The compilation by [148] contains several chapters detailing research on models, tools and techniques and it included case studies in counter-terrorism analysis and open source intelligence. Free text news summaries of the Global Terrorism Database (GTD) was analyzed in [149] to determine incident types. [150] developed a system for analysis of news in real-time to obtain terror incidences, proposing a safety assessment formula based on terror activity of places, and [54] used news and social media content to conduct threat assessment of various airports. The development of a model for the ontological analysis of news [151] used the terrorism domain as a case study and analyzed terrorist news. Statistics: 8 works.

#### 4.11. Disasters/Risks

The multilingual event extraction system in [152] was developed to analyze news aggregated by the European Media Monitor (EMM) system; it extracts violent events as well as natural disaster events. Global news (GDELT-sourced) was analyzed by [153,154] to reveal the structure of coverage of disasters and the determinants of disasters.

Towards supporting aid services during a disaster, news was used by [155] in the research to determine the possibility of/utility of open source data in providing situational awareness for aid agencies to enhance response operations. To enhance speedy and appropriate attention to the needs of the public and responders during a crisis event, [156] developed a system to summarize data streams including news. A similar objective is found in [157] where analytics is used to provide information from news and social networks for emergency management operations.

The extraction of events and the generation of summaries for same from newspaper articles, along with a geographic marker for the event is researched in [158]. In [159] (and extended further in [160]), information sources including news, are mined to obtain demographics, time, location, and summaries.

The system features interactive visualization; a case of gender-based violence and displacement are demonstrated.

To aid journalists in detecting and monitoring disasters, [161] developed a model to automatically identify from local news, the local disaster events which may become global news in a space of 24 hours. The creation of mind maps from news sources was examined in [162] and the idea applied to the disaster domain. In commonality with a number of other works in this section [163] also tackled the extraction of natural disaster occurrences, the end in their case is to populate a store with the information such as disaster type, time, and location extracted from the news. [164] is yet another extraction-focused work utilizing news and social media (Twitter) data to aid disaster management activities. Statistics: 13 works.

#### 4.12. Sustainability/Environment

Newspaper content can assist in the investigation of various issues related to environment and nature. For example, a traditional content analysis of US newspapers on climate change [165]. This research showed the perspectives of various parties involved in the climate change discussion such as 'special interests', scientists and politicians. It further showed the waning and rising of the voices of those parties in the coverage of climate change in the news. The coverage of such environmental issues in the media makes news a suitable data source for various automated algorithms to work with.

News analytics has been applied to the extraction of sustainability indicators; [166] identified, tracked, and reported the indicators of a particular region. By analyzing news [167] produced 58 rules between climate change and food-related vocabularies drawn from news; the study examined the correlation between climate change and food related content. The relationship between water and society in two different countries - Japan and the United States - was studied using news analytics techniques in [168]. A system for the extraction of events from some news of interest was the interest in [158], the author further applied the system to automatic extraction of information about past flood disasters. The social conflicts associated with exploitation of natural resources in four countries was measured using analytics in [169].

To deepen the literature on organized climate scepticism, a corpus of climate-skepticism was created and analysed in [170] to measure key themes in the corpus using analytics methods. [140] showcased a news analytics methodology for the automatic construction of a database of flood events from the data in news reports including causes, and magnitude of the flood. [171] studied the distribution of natural hazards in China by mining news to obtain a natural hazard events dataset, and analyzed it to unveil a spatio-temporal view of natural hazards.

Sustainability is of interest to government, non-government, business, and academia groups; to determine sustainable supply chain, analytics of content from various sources including news was carried out by [127]. Articulate [172], an open source tool for the discovery and management of news coverage on an issue of interest was validated on news about drought and flooding in California and Houston respectively [172]. Statistics: 10 works.

#### 4.13. Information Exploration, Representation, and Visualization

In [3] Pagerank, the mother algorithm of the Google search engine, was referred to as the commonest example of news analytics in use. The rise in volume of data, in particular unstructured data such as news provoked a need to search and explore content in a better manner than obtaining and reading the content personally and in a linear fashion. Search systems are among the earliest applications found in this section. Velthune [173] was developed as an academic news search engine. It retrieves, ranks, indexes, classifies, clusters, and delivers personalized news information extracted from various sources. Lydia [174] is a news analytics system which tracks who is spoken about per time, where the discussions about them emanated, and what is being said about them. [84] automatically generated Newsmap - a knowledge map for exploring intelligence and knowledge captured in business and health news.

The coverage of an entity in the news as a function of a location (of the news source) was used in determining the regional bias of entities in [175]. News stories were analyzed to yield a more informative and limited (in terms of content size) exploration in [176]. SemanticPress, a tool for the generation of press review automatically for the purposes of eGovernment and eParticipation involving the analysis and presentation of news content was presented in [177]. Newsstand [178] leverages the rich geographic content in news to create a system that places news in a map interface. [60] built a system based on the analysis of news for public opinion to enable analysts to consume and analyze news relating to the 2008 US presidential election. In [179] an exploration for news in the terrorist and disaster domains are analyzed on polarity and presented via a visual analytics tool. A visual analytics system for a large collection of news involving polarity, spatial and entity analysis was developed in [180]. Real-time visual analysis of news utilizing threads created from news stories is the subject of [181]. Another visual analytics system described in [182,183], facilitates analysis of similar topics and their evolutions (merging from/splitting into other stories) over time.

While other works in this section focused on visualizing content, [141] analyzed location, and tone or sentiment of news articles and visualized the analysis as heatmaps. In an assistive technology work, [184] simplified news content using an event-centered approach thereby reducing complexity and volume of content to obtain the key information. NewsViews [185], another news visualization system, creates thematic maps from analyzing news content. The AESTHETICS system [186] creates visualizations from the analysis of news content, their categories and entities in the news. The conversion of news articles in qualitative/unstructured form into graphs, tables, keywords, and word vectors is described in [116]. A mapping of news collection into a 2D space illustrating similar news versus dis-similar as a way of giving context to unread news items is described in [187]. In the case of NewsStand [118], a spatio-temporal browser, a map interface allows the query of news items mapped to geo-mentions in them.

The creation of an aggregated news system with thematic grouping and sentiment analysis on real world streams of German news is described in [188]. MediaGist[189] combines summarization, polarity analysis and multilingual aggregation of news to draw the variance in the coverage of the same news story across borders. [121] used the Narrative Visualization technique to allow computers analyze news and present the various sides of the story. They used this to highlight social and economic relationships in financial news. In another re-presentation of news, TimeMachine[190] presents a news search and visualization system which is entity-based, returning related entities for whatever query users present. A temporal aspect is also involved in TimeMachine's processing and so different entity lists may be returned for the same query in a different period. To track the evolution of topics in rapidly changing text data, [191] described an approach to automatically identify topics in incoming documents and match them with an appropriate existing topic immediately preceding them in time. A framework for analyzing civil unrest events from news, with its visualizer, was presented in [31].

The NewsBird system [192,193] is a news aggregator developed to show opposing perspectives on topics in the news. A method to aggregate and organize news datasets into a visualization of major stories is shown in [194]. The creation of an automatic mind map generator to create mind maps from news is the focus in [162]. The concept of semantic relatedness which centers around co-mention of places in news topics is center to the application in [195] which is a computational framework to automatically identify semantic relatedness in news datasets. The work included an exploratory visualization and analysis of geographic distances with respect to semantic relatedness. News from different sources are analyzed in [57] to extract crime related entities and events to create a crime register, a visualization tool allows various interactions with the results from the analysis. Monitoring a news stream for significant co-occurrences of terms with a reference corpus and visualizing the analysis semantically was considered in [196]. The work of [197] focused on the extraction of structured summaries from news by augmenting news with social media tags to derive a "Social Tree", a hierarchical structure of keywords and social tags. Statistics: 34 works.



#### 4.14. Technology

[198] undertook a study of news articles and scientific literature covering the smart services ecosystem and defined the concept of “smart service system” based on the results of the analysis. In [199] the interaction between robotics and social issues in Japan was investigated by pursuing a news analytics study of newspaper articles and academic articles on social robotics. The study highlighted technological solutions available for certain social challenges in Japan. Statistics: 2 works.

#### 4.15. Journalistic Processes, Language Use and Media/News Production

The NSContrast system [200] provided a means of discovering contrary views on matters from different news sources, and characterizing the differences between news sites. The impact of news reports on opioid abuse was investigated by analyzing a dataset of news [87]. An analysis of the coverage of 2007 Kenyan elections in Kenyan newspapers and the UK/US press using computational algorithms was investigated in [66]. Identifying and comparing news frames was the motivation for LingoScope, a visual analytics tool described in [201]. [202] explored a system to automatically predict what the popular news would be, and in [203] news selection for front-page placement was explored to help news aggregators select which articles from a news stream are important. The determinants of the number of comments triggered by a news article was explored in [204]. To assist journalists explore news coverage across countries and across five different languages [189] built MediaGist for cross-lingual analysis of aggregated news and comments.

The application of news analytics techniques to analyze news for large scale analysis of news collections to discover news frames or evaluate news content was explored in [205]. [98] used news analytics to compare topic coverage from articles drawn from two media sources - Twitter and traditional news publications. To allow the discovery of diverse perspectives on the same topic, the NewsBird aggregator was developed in [192,193]. How news volume corresponds to disease incidence rates was investigated in [100]. [126] studied South Korean media to understand the differences between how journalists presented gambling and how the public perceives the news.

The creation of a news dataset and a feature data set for inquiry into news production was undertaken in [77]. An approach to authenticate journalists by cross checking whether their statements of what would hold in the future actually happened, called automatic prediction validations (APV) was proposed and built in [137]. Techniques journalists can employ to analyze news and discover insight they can use to steer their own work are explored in [206]. The utility of news analytics techniques towards studying the depiction of migrants and minorities in Korean news was the subject of [207]. The MediaRank, a computational approach to undertake automatic ranking of news sources around the world, based on four journalistic values associated with quality of reports is described in [208]. In [209], a series of means were proposed towards the end of predicting evergreen posts. Statistics: 20 works.

#### 4.16. Notoriety

The problem of predicting whether a news article will generate a lot of uncivil content in its comments section was tackled in [80], and the [210] characterized hate speech content against immigrants in Korea by analyzing news comments. Statistics: 2 works.

#### 4.17. Gender Issues

[211] showed the relationship between topic of news articles and gender bias by using news analytics. Although the objective of the research by [47] was not directly targeted at gender, the demonstration of their analytics method on news articles revealed how women and children were mostly the victims of Crime against the Person. Statistics: 2 works.

#### 4.18. Gambling

The dynamics of news reports and public interest involving the gambling industry in South Korea was studied in [126], covering various dimensions of the topic. Statistics: 1 work.

#### 4.19. Migration

Towards border security, the work in [43] presented a real-time event extraction focused on border-related acts including illegal migration, from news reports in Europe. To evaluate their work aimed at the automatic discovery of framing of latent personas from documents, [212] analyzed a collection of immigration news. The detection of forced migration via a targeted event detection approach was discussed in [213]; the same problem was the subject of the thesis [214] which mined news articles to predict forced migration. To understand the public-migrant interaction on a smaller scale, [210] analyzed news comments to characterize news comments for hate speech against foreign migrant workers. The idea to integrate traditional indicators with news and social media analytics to predict migration and the attendant migration pattern was proposed in [215], and [207] explored the application of analytics to large corpus to investigate the representation of minorities in news reporting. Statistics: 7 works.

#### 4.20. Sports

Observing the fact that studying the lifestyle of triathlon athletes is of interest to social sport scientists, [216] analyzed triathlon athletes blogs and news feeds from athletes' websites. While the authors did not use mainstream news sources, their objective may well be achieved with a switch to sports news covering the triathlon niche. Statistics: 1 work.

#### 4.21. Societal/Government

News analytics was used in exploring societal interests in [217]. In [175] what entities dominate local discussions across geographical locations is the interest, and these are extracted from various regional news sources. The NSContrast system proposed in [200] allows the exploration of the interests of a particular country. Historical references in news is studied in [138] who analyzed news collection from different countries to reveal hidden correlations. [218] explored news analytics towards understanding the various macro-level changes in the society based on the tone of news content and geographic analysis of news content. The generation of fine-grained heatmaps corresponding to the emotions of people at a location as determined from the content from various text sources including news is the subject of [141].

In [219], the GDELT database coverage of events in Singapore as drawn from a variety of sources was tested for quality and accuracy. The discussion about water and the society in the news was the intent in [168]. Measuring the strength of nations or their influence according to media reports or events in the news was considered in [220] which analyzed the GDELT data corresponding to USA, Russia, and China activities in the news. Another cross-country work mined news to determine social conflicts arising as a result of mining activities in Australia, Canada, Chile, and Peru [169]. The utilization of the newspaper industry in China as both a government mouthpiece and also as a market and economy information provider is studied in [74] via textual analysis of a news corpus. MediaGist [189] presented a tool that allowed the detection of controversial topics and their exploration from a corpus of international news, similarly, [221] analyzed aggregated news to extract interesting topics. The system was used to compare news articles from different regions.

The co-occurrence of mentions of city names in news topics can be considered semantic relatedness as noted in [195], where the concept was automatically examined in a large news corpus. An analysis of the Arabic news collection to reveal top topics across the corpus and the coverage scope of the 24 different Arabic newspapers examined was reported in [222]. The extraction of key noun terms of social-problem-related topics in the news is studied in [223]. Statistics: 16 works.

#### 4.22. Emerging/Interesting Entities and Events

A named entity recognition system in newspapers that uses the distribution of words in news articles to determine named entities was described in [224]. The Lydia system described in [174] extracts a relational model of entities in the news, it could determine what entities are the subject of the news at a particular place. In likeness to the Lydia system, [225] presented their approach that mines a

large collection of news to extract topics, trends and linked entities. Another automatic analysis work [226], extracted relationships between persons in the news and populated these on maps. [227] did not just extract entities but also extracted significant events from daily news. A multilingual news mining for extracting who, where, and what is reported in the news was presented in [228], and a technique for the extraction of social networks of actors from text was demonstrated in [61]. [229]'s interest was tracking of the set of actions or activities culminating in a major event or news story, the author's interest was alike to [180] which examined news from different countries considering the key topics, tone, and the quantitative evolution of entities mentioned in the news in a temporal manner.

The extraction of events from Vietnamese news and an online event monitoring system was the subject of in [230]. The presentation of contrasting news views using an automated technique was the contribution in [231] while a [68]'s contribution was an approach to analyze news to quickly extract key actors and their networks in the context of an emerging or current major event such as the Arab Spring of 2011. [232] investigated the problem of determining the day's key news story for users, given streams of articles from several sources, while [233] worked on discovering and semantically typing emerging entities in news. Extracting organization name from business news in Russian language received attention in [234], The system in [235] processed a multilingual news collection to extract the major news from the collection and to track related news over a period of seven days. Visualizing entity trends extracted via news analytics was treated in [186], while in [185] concepts and locations are mined from news and used in the creation of thematic visualizations. Certain entities in text may be ambiguous and their discovery is difficult, the efforts in [236] was towards discovering and resolving ambiguous named entities. The use of named entities in ranking news articles was explored in [237]. Integrated management and mining of multimodal content was described for the news and events domain in [238], while in [239] a method for easy exploration of key news events and summary related information was discussed.

News events are linked by the semantic relations they possess to build an "event centrality management system" in [92]. The system described in [240] is also similar in extracting topics and discovering the event-relationships. Named entity recognition in crime news was considered in [241], business-critical events were analyzed in [242] and Chinese news headlines analyzed in [243] to detect and follow trending events. News analytics techniques were applied in [244] to study the newsflow theory, and in [245] socio-economic indicators like food prices were predicted by analyzing news events. News and allied sources were analyzed and formed the data for GeoSentiment [246] - a visualization tool for event-triggered sentiments across location. The use of automatically identified entities and relations between them to drive news exploration and visualization was the highlight in [190]. The construction of knowledge graphs based on news events is discussed in [247], the authors noted the benefits of event-centric news exploration. Event detection and linking to construct stories in large news streams was presented in [248]. A method based on analyzing events in news stories to suggest news that may be used as external sources for Wikipedia articles is proposed in [249]. Statistics: 34 works.

#### 4.23. Negotiation

News analytics has been used in research considering the provision of valuable background information to negotiators from news articles [110]. Statistics: 1 work.

#### 4.24. Education

News articles have been analyzed to understand the content simplification process. In [250] original and abridged news articles were compared using various algorithms. To measure science literacy in the society, [251] analyzed a concept map drawn from a large corpus of Chinese news articles. The media's coverage of MOOCs Massive Open Online Course(s) was examined in [252] using news analytics techniques to understand public discourse of MOOCs. Statistics: 3 works.

#### 4.25. Miscellaneous Applications

We identified some other applications of news analytics which do not identify as a major theme or category within the general scheme of things in our real world; the works are cataloged in this section. The discovery of the influences of peak news on other news topics and how it exposes knowledge about interests of the society and society's behavior received attention in [253]. The analysis in [254] was concerned with the clustering of face images from news content and labeling them appropriately. A technique for the extraction of named entities by using the distribution of words in news content was presented in [224], NewsExplorer combined various text analysis tools across 19 languages, creating stories from related news; extracting and aggregating from these multilingual collections and linking them all. A set of transformations to model actors and their interactions in news stories was proposed in [176], and [255] presented a technique for the automatic population of ontology from news. To analyze a volume of news and present news efficiently and meaningfully, incident threading was explored in [256], similarly the mining and ranking of stories presented within similar time windows was treated in [257]. Contention in news reports was investigated in [231], the authors classified news reports automatically based on disputant relations.

The Global Data on Events, Location and Tone (GDELT), a dataset of geolocated events with global coverage is described in [258] with respect to its news sources and processing pipeline among others. We note that the GDELT dataset has been used in several news analytics research we cover in this work. Extraction of new entities emerging in news for updating an ontology is described by [233]. [259] analyzed news streams, connects relevant topics with summarized Twitter contents around it including relevant hashtags. A research with a unique direction, [219], analyzed GDELT data on Singapore events to determine its completeness and validity. In [260] segments of news content that are argument elements were identified. The news analysis and mining task is studied from a link-eccentric perspective in [240], including the process for discovering topics and linking entities in events. Questions about online news characteristics such as which was the most relevant or what is the lifespan of news have been studied, using a set of top Italian news sources and agencies [261]. An automated attempt at producing press reviews is discussed in [262], and in [263] the problem of comparing news in different languages and linking articles covering similar events but in different languages was examined. While we have centered on analyzing textual content to derive some form of quantity which may be mined, research to do the opposite can be found. [264] discussed the feasibility and process of discovering interesting observations from structured data and producing news from it automatically. Also in the same vein is the research on Reuters Tracer, an end-to-end system for production of news from Twitter data [265]. To understand the creative and cultural spheres across China, [266] undertook a news analytics exploration of the news. In their experiments, [249] presented an automated means of predicting candidate news for Wikipedia external references from a daily news stream. Statistics: 21 works.

## 5. Discussions

There is no doubt that there is an abundance of news analytics research works outside the presented collection, neither is the set of categories we present here an exhaustive one or a perfect grouping. Within the scope of works we have presented, a few observations can be made.

### 5.1. On the Datasets Used in Examined Works

- We observed a number of research interests around live news and real-time news analytics. However, since these are different with regards to approaches involved as compared to those in the works we consider, they are not included in this exploration. We intend to carry out a separate examination of these in future work.
- Comments on news articles, which are a sort of derivative of news articles in our view, are also used in news analytics tasks. Comments datasets may be used solely, or used in conjunction with the news articles.

- Social media text, often from Twitter, features as a common augmentation of news data used in analytics.
- Data from GDELT is also used. However most researchers obtain their datasets directly from raw news than from a pre-compiled source such as Global Data on Events, Location and Tone (GDELT). The Europe Media Monitor (EMM), is another source of pre-compiled data from news that is identified in our review.

### 5.2. On Focal Interests of the Works

- The purpose of news analytics is not always predictive, as we discovered in our review of existing literature in this field. Some authors analyzed news content with the intent of
  - i) producing concise reports,or
  - ii) forecasting,
  - iii) building an information tool,
  - iv) providing a better interface for discovering and consuming the issues in the news,
  - v) deriving concise reports from news or allowing users to get a better scope of present issues, among several others.

### 5.3. On Reach or Geographical Scope of News Analytics

Almost all the continents have been the subject of one news analytics research or the other The presence of major news analytics projects from consortia and government in Europe, North America, and also in Australia/Asia Pacific region is also noted.

## 6. Conclusion

Rather than attempting to be exhaustive or comprehensive, we are motivated by an inquisitiveness to explore what range of non-financial news analytics works exist within our literature search; i.e. given a representative set of published works on news analytics, what may we find? The question succinctly captures what we sought to do, and answering that question is what we have done. We have identified domains in the society that has been observed or investigated through news analytics.

Our survey show that news analytics is diverse, serves various purposes, and features interesting applications and usages. Beyond financial news analytics with popularity which can not be disputed in the overall field, news analytics for security intents (comprising the crime, unrest, protest, terrorism, disasters and violence themes) is another popular area. The security theme spans five (5) of the 25 categories we covered; by this, security is the predominant news analytics research sub-field after the markets. The importance of the media and the impact on the society has been researched over time, and now news is finding utility in a variety of big data applications bordering on security in the society. Everyday, millions of events are covered in the news; government and law enforcement operatives are interested in these news; they want to monitor, track, and analyze this information. This could be one reason why there is much interest in analyzing news for security issues. Furthermore, security concerns seem to be generally increasing both in the populace and in government. We believe that the security sub-field will witness more applications and results as machine learning and artificial intelligence algorithms improve, giving more power to the text mining and natural language processes employed in news analytics systems.

Social science researchers are also picking up on text mining as a qualitative research methodology, and substantial social computing research deals with the analysis of text, sentiments, and social networks. In this survey we have works that mine news to understudy the social network of important figures [61], [123]. This is of course different from mining social network connections in a traditional social network platform setting, e.g. Facebook. Despite the source of data and the presentation of the data source being different, the same network analysis algorithms and approaches used in mining social networks can be applied in mining the social connection graph obtained from news analytics. Applications that extract social networks from news content, whether in real-time or

from news collections, can further combine with other news analytics intents or processes to yield a multidimensional or multi-faceted analysis and understanding of news text collections.

## References

1. Mayring, P. Qualitative content analysis. *A companion to qualitative research* **2004**, *1*, 159–176.
2. Neuendorf, K.A.; Kumar, A. Content analysis. *The international encyclopedia of political communication* **2015**, pp. 1–10.
3. Das, S.R. News analytics: Framework, techniques and metrics. *The Handbook of News Analytics in Finance* **2011**, *2*.
4. Martín-Martín, A.; Orduna-Malea, E.; Thelwall, M.; López-Cózar, E.D. Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics* **2018**, *12*, 1160–1177.
5. Gusenbauer, M. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics* **2019**, *118*, 177–214. <https://doi.org/10.1007/s11192-018-2958-5>.
6. Feldman, R.; Sanger, J.; et al. *The text mining handbook: advanced approaches in analyzing unstructured data*; Cambridge university press, 2007.
7. Gupta, V.; Lehal, G.S.; et al. A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence* **2009**, *1*, 60–76.
8. Glaser, B.; Strauss, A. The discovery of grounded theory. 1967. *Weidenfeld & Nicolson, London* **1967**, pp. 1–19.
9. Corbin, J.M.; Strauss, A. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology* **1990**, *13*, 3–21.
10. Batty, C.D. *An introduction to the Dewey decimal classification*; Asia Publishing House (1966), 2017.
11. Banerjee, A.; Bandyopadhyay, T.; Acharya, P. Data analytics: Hyped up aspirations or true potential? *Vikalpa* **2013**, *38*, 1–12.
12. Davenport, T.H. The rise of automated analytics. *The Wall Street Journal* **2015**.
13. Kent, E.L. Text analytics—techniques, language and opportunity. *Business Information Review* **2014**, *31*, 50–53.
14. Mitra, L.; Mitra, G. Applications of news analytics in finance: A review. *The handbook of news analytics in finance* **2011**, 596.
15. Cambria, E.; Rajagopal, D.; Olsher, D.; Das, D. Big Social Data Analysis. In *Big Data Computing*; Akerkar, R., Ed.; Chapman and Hall/CRC, 2013; pp. 401–414. <https://doi.org/10.1201/b16014-19>.
16. Ball, P. News mining might have predicted Arab Spring. *Nature* **2011**. <https://doi.org/10.1038/news.2011.532>.
17. Gao, J.; Leetaru, K.H.; Hu, J.; Cioffi-Revilla, C.; Schrodt, P. Massive Media Event Data Analysis to Assess World-Wide Political Conflict and Instability. In *Social Computing, Behavioral-Cultural Modeling and Prediction*; Hutchison, D.; Kanade, T.; Kittler, J.; Kleinberg, J.M.; Mattern, F.; Mitchell, J.C.; Naor, M.; Nierstrasz, O.; Pandu Rangan, C.; Steffen, B.; et al., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; Vol. 7812, pp. 284–292. [https://doi.org/10.1007/978-3-642-37210-0\\_31](https://doi.org/10.1007/978-3-642-37210-0_31).
18. Hua, T.; Lu, C.; Ramakrishnan, N.; Chen, F.; Arredondo, J.; Mares, D.; Summers, K. Analyzing Civil Unrest through Social Media. *Computer* **2013**, *46*, 80–84.
19. Keneshloo, Y.; Cadena, J.; Korkmaz, G.; Ramakrishnan, N. Detecting and forecasting domestic political crises: a graph-based approach. In *Proceedings of the Proceedings of the 2014 ACM conference on Web science - WebSci '14*, Bloomington, Indiana, USA, 2014; pp. 192–196. <https://doi.org/10.1145/2615569.2615698>.
20. Muthiah, S. Forecasting protests by detecting future time mentions in news and social media. PhD thesis, Virginia Tech, 2014.
21. Ramakrishnan, N.; Korkmaz, G.; Kuhlman, C.; Marathe, A.; Zhao, L.; Hua, T.; Chen, F.; Lu, C.T.; Huang, B.; Srinivasan, A.; et al. 'Beating the news' with EMBERS: forecasting civil unrest using open source indicators. In *Proceedings of the Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*, New York, New York, USA, 2014; pp. 1799–1808. <https://doi.org/10.1145/2623330.2623373>.
22. Doyle, A.; Katz, G.; Summers, K.; Ackermann, C.; Zavorin, I.; Lim, Z.; Muthiah, S.; Zhao, L.; Lu, C.T.; Butler, P.; et al. The EMBERS architecture for streaming predictive analytics. In *Proceedings of the 2014 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, 2014; pp. 11–13. <https://doi.org/10.1109/BigData.2014.7004477>.

23. Doyle, A.; Katz, G.; Summers, K.; Ackermann, C.; Zavorin, I.; Lim, Z.; Muthiah, S.; Butler, P.; Self, N.; Zhao, L.; et al. Forecasting Significant Societal Events Using The Embers Streaming Predictive Analytics System. *Big Data* **2014**, *2*, 185–195. <https://doi.org/10.1089/big.2014.0046>.
24. Korkmaz, G.; Cadena, J.; Kuhlman, C.J.; Marathe, A.; Vullikanti, A.; Ramakrishnan, N. Combining Heterogeneous Data Sources for Civil Unrest Forecasting. In Proceedings of the Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 - ASONAM '15, Paris, France, 2015; pp. 258–265. <https://doi.org/10.1145/2808797.2808847>.
25. Muthiah, S.; Huang, B.; Arredondo, J.; Mares, D.; Getoor, L.; Katz, G.; Ramakrishnan, N. Planned Protest Modeling in News and Social Media. In Proceedings of the Twenty-Seventh IAAI Conference, 2015.
26. Cadena, J.; Korkmaz, G.; Kuhlman, C.J.; Marathe, A.; Ramakrishnan, N.; Vullikanti, A. Forecasting Social Unrest Using Activity Cascades. *PLOS ONE* **2015**, *10*, e0128879. <https://doi.org/10.1371/journal.pone.0128879>.
27. Ramakrishnan, N.; Lu, C.T.; Marathe, M.; Marathe, A.; Vullikanti, A.; Eubank, S.; Leman, S.; Roan, M.; Brownstein, J.S.; Summers, K.; et al. Model-Based Forecasting of Significant Societal Events. *IEEE Intelligent Systems* **2015**, *30*, 86–90. <https://doi.org/10.1109/MIS.2015.74>.
28. Saraf, P.; Ramakrishnan, N. EMBERS AutoGSR: Automated Coding of Civil Unrest Events. In Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, San Francisco, California, USA, 2016; pp. 599–608. <https://doi.org/10.1145/2939672.2939737>.
29. Wang, W.; Ning, Y.; Rangwala, H.; Ramakrishnan, N. A Multiple Instance Learning Framework for Identifying Key Sentences and Detecting Events. In Proceedings of the Proceedings of the 25th ACM International on Conference on Information and Knowledge Management - CIKM '16, Indianapolis, Indiana, USA, 2016; pp. 509–518. <https://doi.org/10.1145/2983323.2983821>.
30. Muthiah, S.; Huang, B.; Arredondo, J.; Mares, D.; Getoor, L.; Katz, G.; Ramakrishnan, N. Capturing Planned Protests from Open Source Indicators. *AI Magazine* **2016**, *37*, 63. <https://doi.org/10.1609/aimag.v37i2.2631>.
31. Saraf, P.; Self, N.; Ramakrishnan, N. Who, When, Where and Why? Visualizing Civil Unrest Events. p. 4.
32. Ning, Y.; Muthiah, S.; Rangwala, H.; Ramakrishnan, N. Modeling Precursors for Event Forecasting via Nested Multi-Instance Learning. In Proceedings of the Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1095–1104.
33. Korkmaz, G.; Cadena, J.; Kuhlman, C.J.; Marathe, A.; Vullikanti, A.; Ramakrishnan, N. Multi-source models for civil unrest forecasting. *Social Network Analysis and Mining* **2016**, *6*, 50. <https://doi.org/10.1007/s13278-016-0355-8>.
34. Qiao, F.; Li, P.; Deng, J.; Ding, Z.; Wang, H. Graph-Based Method for Detecting Occupy Protest Events Using GDELT Dataset. In Proceedings of the 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, Xi'an, China, 2015; pp. 164–168. <https://doi.org/10.1109/CyberC.2015.77>.
35. Papanikolaou, K.; Papageorgiou, H.; Pappasantopoulos, N.; Stathopoulou, T.; Papastefanatos, G. "Just the Facts" with PALOMAR: Detecting Protest Events in Media Outlets and Twitter. In Proceedings of the Tenth International AAAI Conference on Web and Social Media, 2016.
36. Atkinson, M.; Piskorski, J.; Tanev, H.; Zavarella, V. On the Creation of a Security-Related Event Corpus. In Proceedings of the Proceedings of the Events and Stories in the News Workshop, Vancouver, Canada, 2017; pp. 59–65. <https://doi.org/10.18653/v1/W17-2709>.
37. Kang, W.; Chen, J.; Li, J.; Liu, J.; Liu, L.; Osborne, G.; Lothian, N.; Cooper, B.; Moschou, T.; Neale, G. Carbon: Forecasting Civil Unrest Events by Monitoring News and Social Media. In *Advanced Data Mining and Applications*; Cong, G.; Peng, W.C.; Zhang, W.E.; Li, C.; Sun, A., Eds.; Springer International Publishing: Cham, 2017; Vol. 10604, pp. 859–865. [https://doi.org/10.1007/978-3-319-69179-4\\_62](https://doi.org/10.1007/978-3-319-69179-4_62).
38. Osborne, G.; Lothian, N.; Neale, G.; Moscou, T.; Nguyen, A.; Chen, J.; Kang, W.; Cooper, B.; et al. The Beat the News System: Forecasting Social Disruption via Modelling of Online Behaviours. *Journal of the Australian Institute of Professional Intelligence Officers* **2017**, *25*, 35.
39. Sharma, K.; Sehgal, G.; Gupta, B.; Sharma, G.; Chatterjee, A.; Chakraborti, A.; Shroff, G. A complex network analysis of ethnic conflicts and human rights violations. *Scientific Reports* **2017**, *7*, 8283. <https://doi.org/10.1038/s41598-017-09101-8>.
40. Ning, Y.; Muthiah, S.; Ramakrishnan, N.; Rangwala, H.; Mares, D. When do Crowds Turn Violent? Uncovering Triggers from Media. In Proceedings of the 2018 IEEE/ACM International Conference on

- Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, 2018; pp. 77–82. <https://doi.org/10.1109/ASONAM.2018.8508280>.
41. Thenmozhi, D.; Aravindan, C.; Shyamsunder, A.; Vishwanathan, A.; Pujari, A.K. Extracting Protests from News using LSTM Models with different Attention Mechanisms. In Proceedings of the CLEF (Working Notes), 2019.
  42. Dozier, C.; Jackson, P. Mining Text for Expert Witnesses. *IEEE Software* **2005**, *22*, 94–100. <https://doi.org/10.1109/MS.2005.70>.
  43. Piskorski, J.; Atkinson, M.; Belyaeva, J.; Zavarella, V.; Huttunen, S.; Yangarber, R. Real-time text mining in multilingual news for the creation of a pre-frontier intelligence picture. In Proceedings of the ACM SIGKDD Workshop on Intelligence and Security Informatics - ISI-KDD '10, Washington, D.C., 2010; pp. 1–9. <https://doi.org/10.1145/1938606.1938610>.
  44. Piskorski, J.; Atkinson, M. Frontex real-time news event extraction framework. In Proceedings of the Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11, San Diego, California, USA, 2011; p. 749. <https://doi.org/10.1145/2020408.2020527>.
  45. Alruily, M.; Ayes, A.; Al-Marghilani, A. Using Self Organizing Map to cluster Arabic crime documents. In Proceedings of the Proceedings of the International Multiconference on Computer Science and Information Technology, Wisla, 2010; pp. 357–363. <https://doi.org/10.1109/IMCSIT.2010.5679616>.
  46. Mohamad Ali, N.; Mohd, M.; Lee, H.; Smeaton, A.F.; Crestani, F.; Mohd Noah, S.A. i-JEN: Visual Interactive Malaysia Crime News Retrieval System. In *Visual Informatics: Sustaining Research and Innovations*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; Vol. 7067, pp. 284–294. [https://doi.org/10.1007/978-3-642-25200-6\\_27](https://doi.org/10.1007/978-3-642-25200-6_27).
  47. Sudhahar, S.; Franzosi, R.; Cristianini, N. Automating quantitative narrative analysis of news data. In Proceedings of the Proceedings of the Second Workshop on Applications of Pattern Analysis, 2011, pp. 63–71.
  48. Tseng, Y.H.; Ho, Z.P.; Yang, K.S.; Chen, C.C. Mining term networks from text collections for crime investigation. *Expert Systems with Applications* **2012**, *39*, 10082–10090. <https://doi.org/10.1016/j.eswa.2012.02.052>.
  49. Bsoul, Q.; Salim, J.; Zakaria, L.Q. An Intelligent Document Clustering Approach to Detect Crime Patterns. *Procedia Technology* **2013**, *11*, 1181–1187. <https://doi.org/10.1016/j.protcy.2013.12.311>.
  50. Alruily, M.; Ayes, A.; Zedan, H. Crime profiling for the Arabic language using computational linguistic techniques. *Information Processing & Management* **2014**, *50*, 315–341. <https://doi.org/10.1016/j.ipm.2013.09.001>.
  51. Mookiah, L.; Eberle, W.; Siraj, A. Survey of Crime Analysis and Prediction. In Proceedings of the FLAIRS Conference, 2015, pp. 440–443.
  52. Jayaweera, I.; Sajeewa, C.; Liyanage, S.; Wijewardane, T.; Perera, I.; Wijayasiri, A. Crime analytics: Analysis of crimes through newspaper articles. In Proceedings of the 2015 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 2015; pp. 277–282. <https://doi.org/10.1109/MERCon.2015.7112359>.
  53. Das, S.; Choudhury, M.R. A geo-statistical approach for crime hot spot prediction. *International Journal of Criminology and Sociological Theory* **2016**, *9*.
  54. Khandpur, R.; Ji, T.; Ning, Y.; Zhao, L.; Lu, C.T.; Smith, E.; Adams, C.; Ramakrishnan, N. Determining Relative Airport Threats from News and Social Media. In Proceedings of the Innovative Applications of Artificial Intelligence, 2017.
  55. Matto, G.; Mwangoka, J. Detecting crime patterns from Swahili newspapers using text mining **2017**. p. 12.
  56. Nokhbeh Zaeem, R.; Manoharan, M.; Yang, Y.; Barber, K.S. Modeling and analysis of identity threat behaviors through text mining of identity theft stories. *Computers & Security* **2017**, *65*, 50–63. <https://doi.org/10.1016/j.cose.2016.11.002>.
  57. Dasgupta, T.; Dey, L.; Saha, R.; Naskar, A. Automatic Curation and Visualization of Crime Related Information from Incrementally Crawled Multi-source News Reports. In Proceedings of the Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations, 2018, pp. 103–107.
  58. Po, L.; Rollo, F. Building an Urban Theft Map by Analyzing Newspaper Crime Reports. In Proceedings of the 2018 13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), Zaragoza, 2018; pp. 13–18. <https://doi.org/10.1109/SMAP.2018.8501866>.
  59. Lerman, K.; Gilder, A.; Dredze, M.; Pereira, F. Reading the markets: forecasting public opinion of political candidates by news analysis. In Proceedings of the Proceedings of the 22nd International Conference on Computational Linguistics - COLING '08, Manchester, United Kingdom, 2008; Vol. 1, pp. 473–480. <https://doi.org/10.3115/1599081.1599141>.



60. Wanner, F.; Rohrdantz, C.; Mansmann, F.; Oelke, D.; Keim, D.A. Visual Sentiment Analysis of RSS News Feeds Featuring the US Presidential Election in 2008 **2009**. p. 8.
61. Danowski, J.A.; Cepela, N.T. Automatic Mapping of Social Networks of Actors from Text Corpora: Time Series Analysis. In Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining, Athens, Greece, 2009; pp. 137–142. <https://doi.org/10.1109/ASONAM.2009.71>.
62. Davulcu, H.; Ahmed, S.T.; Gokalp, S.; Temkit, M.H.; Taylor, T.; Woodward, M.; Amin, A. Analyzing Sentiment Markers Describing Radical and Counter-Radical Elements in Online News. In Proceedings of the 2010 IEEE Second International Conference on Social Computing, Minneapolis, MN, USA, 2010; pp. 335–340. <https://doi.org/10.1109/SocialCom.2010.55>.
63. Park, S.; Ko, M.; Kim, J.; Liu, Y.; Song, J. The politics of comments: predicting political orientation of news stories with commenters' sentiment patterns. In Proceedings of the Proceedings of the ACM 2011 conference on Computer supported cooperative work - CSCW '11, Hangzhou, China, 2011; p. 113. <https://doi.org/10.1145/1958824.1958842>.
64. Schrod, P.A. Automated Production of High-Volume, Near-Real-Time Political Event Data. In Proceedings of the APSA 2010 Annual Meeting Paper, 2010.
65. Neri, F.; Aliprandi, C.; Camillo, F. Mining the web to monitor the political consensus. In *Counterterrorism and Open Source Intelligence*; Springer, 2011; pp. 391–412.
66. Pollak, S.; Coesemans, R.; Daelemans, W.; Lavrač, N. Detecting contrast patterns in newspaper articles by combining discourse analysis and text mining. *Pragmatics* **2011**, *21*, 647–683. <https://doi.org/10.1075/prag.21.4.07pol>.
67. Balasubramanyan, R.; Cohen, W.W.; Pierce, D.; Redlawsk, D.P. Modeling Polarizing Topics: When Do Different Political Communities Respond Differently to the Same News? **2012**. p. 8.
68. Pfeffer, J.; Carley, K.M. Rapid modeling and analyzing networks extracted from pre-structured news articles. *Computational and Mathematical Organization Theory* **2012**, *18*, 280–299. <https://doi.org/10.1007/s10588-012-9122-1>.
69. De Fortuny, E.J.; De Smedt, T.; Martens, D.; Daelemans, W. Media coverage in times of political crisis: A text mining approach. *Expert Systems with Applications* **2012**, *39*, 11616–11622.
70. Grimmer, J.; Stewart, B.M. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* **2013**, *21*, 267–297. <https://doi.org/10.1093/pan/mps028>.
71. Soelistio, Y.E.; Surendra, M.R.S. Simple text mining for sentiment analysis of political figure using naive bayes classifier method. *arXiv preprint arXiv:1508.05163* **2015**.
72. Schrod, P.A.; Analytics, P. Comparing Methods for Generating Large Scale Political Event Data Sets. In Proceedings of the Text as Data meetings, New York University, 16–17, 2015, 2015, pp. 1–32.
73. Bakken, P.F.; Bratlie, T.A.; Marco, C.; Gulla, J. Political News Sentiment Analysis for Under-resourced Languages. In Proceedings of the COLING, 2016.
74. Piotroski, J.D.; Wong, T.; Zhang, T. Political Bias of Corporate News: Role of Conglomeration Reform in China. *The Journal of Law and Economics* **2017**, *60*, 173–207.
75. Yuan, Y. Modeling Inter-country Connection from Geotagged News Reports: A Time-Series Analysis. In *Data Mining and Big Data*; Tan, Y.; Takagi, H.; Shi, Y., Eds.; Springer International Publishing: Cham, 2017; Vol. 10387, pp. 183–190. [https://doi.org/10.1007/978-3-319-61845-6\\_19](https://doi.org/10.1007/978-3-319-61845-6_19).
76. Lazaridou, K.; Krestel, R.; Naumann, F. Identifying Media Bias by Analyzing Reported Speech. In Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA, 2017; pp. 943–948. <https://doi.org/10.1109/ICDM.2017.119>.
77. Horne, B.D.; Dron, W.; Khedr, S.; Adali, S. Sampling the news producers: A large news and feature data set for the study of the complex media landscape. *arXiv preprint arXiv:1803.10124* **2018**.
78. Mueller, H.; Rauh, C. Reading Between the Lines: Prediction of Political Violence Using Newspaper Text. *American Political Science Review* **2018**, *112*, 358–375. <https://doi.org/10.1017/S0003055417000570>.
79. Fang, P.; Ma, M.I.; Gao, J.b. Dynamic Evolution of Multinational Relation's Network in the South China Sea Arbitration Based on Massive Media Data Analysis. *DEStech Transactions on Computer Science and Engineering* **2018**. <https://doi.org/10.12783/dtcse/pcomm2018/23735>.
80. Magu, R.; Hossain, N.; Kautz, H. Analyzing uncivil speech provocation and implicit topics in online political news. *arXiv preprint arXiv:1807.10882* **2018**.
81. Marozzo, F.; Bessi, A. Analyzing polarization of social media users and news sites during political campaigns. *Social Network Analysis and Mining* **2018**, *8*, 1. <https://doi.org/10.1007/s13278-017-0479-5>.

82. Lawrence, R.G. Framing Obesity: The Evolution of News Discourse on a Public Health Issue. *Harvard International Journal of Press/Politics* **2004**, *9*, 56–75. <https://doi.org/10.1177/1081180X04266581>.
83. Pouliquen, B.; Steinberger, R.; Ignat, C.; Käsper, E.; Temnikova, I. Multilingual and cross-lingual news topic tracking. In Proceedings of the Proceedings of the 20th international conference on Computational Linguistics - COLING '04, Geneva, Switzerland, 2004; pp. 959–es. <https://doi.org/10.3115/1220355.1220493>.
84. Ong, T.H.; Chen, H.; Sung, W.k.; Zhu, B. Newsmap: a knowledge map for online news. *Decision Support Systems* **2005**, *39*, 583–597. <https://doi.org/10.1016/j.dss.2004.03.008>.
85. Takahashi, Y.; Miyaki, K.; Nakayama, T. Analysis of news of the Japanese asbestos panic: a supposedly resolved issue that turned out to be a time bomb. *Journal of Public Health* **2007**, *29*, 62–69. <https://doi.org/10.1093/pubmed/fdl081>.
86. Zhang, Y.; Dang, Y.; Chen, H.; Thurmond, M.; Larson, C. Automatic online news monitoring and classification for syndromic surveillance. *Decision Support Systems* **2009**, *47*, 508–517. <https://doi.org/10.1016/j.dss.2009.04.016>.
87. Dasgupta, N.; Mandl, K.D.; Brownstein, J.S. Breaking the News or Fueling the Epidemic? Temporal Association between News Media Report Volume and Opioid-Related Mortality. *PLoS ONE* **2009**, *4*, e7758. <https://doi.org/10.1371/journal.pone.0007758>.
88. Collier, N. What's unusual in online disease outbreak news? *Journal of biomedical semantics* **2010**, *1*, 2.
89. Bertini, E.; Buchmuller, J.; Fischer, F.; Huber, S.; Lindemeier, T.; Maass, F.; Mansmann, F.; Ramm, T.; Regenscheit, M.; Rohrdantz, C.; et al. Visual analytics of terrorist activities related to epidemics. In Proceedings of the 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), Providence, RI, USA, 2011; pp. 329–330. <https://doi.org/10.1109/VAST.2011.6102498>.
90. Lejeune, G.; Brixtel, R.; Doucet, A.; Lucas, N. DANIEL: Language Independent Character-Based News Surveillance. In *Advances in Natural Language Processing*; Hutchison, D.; Kanade, T.; Kittler, J.; Kleinberg, J.M.; Mattern, F.; Mitchell, J.C.; Naor, M.; Nierstrasz, O.; Pandu Rangan, C.; Steffen, B.; et al., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; Vol. 7614, pp. 64–75. [https://doi.org/10.1007/978-3-642-33983-7\\_7](https://doi.org/10.1007/978-3-642-33983-7_7).
91. Lupan, D.; Dascălu, M.; Trăuşan-Matu, Ş.; Dessus, P. Analyzing Emotional States Induced by News Articles with Latent Semantic Analysis. In *Artificial Intelligence: Methodology, Systems, and Applications*; Hutchison, D.; Kanade, T.; Kittler, J.; Kleinberg, J.M.; Mattern, F.; Mitchell, J.C.; Naor, M.; Nierstrasz, O.; Pandu Rangan, C.; Steffen, B.; et al., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; Vol. 7557, pp. 59–68. [https://doi.org/10.1007/978-3-642-33185-5\\_7](https://doi.org/10.1007/978-3-642-33185-5_7).
92. Xu, Z.; Wei, X.; Luo, X.; Liu, Y.; Mei, L.; Hu, C.; Chen, L. Knowle: A semantic link network based system for organizing large scale online news events. *Future Generation Computer Systems* **2015**, *43-44*, 40–50. <https://doi.org/10.1016/j.future.2014.04.002>.
93. Mollema, L.; Harmsen, I.A.; Broekhuizen, E.; Clijnk, R.; De Melker, H.; Paulussen, T.; Kok, G.; Ruiter, R.; Das, E. Disease Detection or Public Opinion Reflection? Content Analysis of Tweets, Other Social Media, and Online Newspapers During the Measles Outbreak in the Netherlands in 2013. *Journal of Medical Internet Research* **2015**, *17*, e128. <https://doi.org/10.2196/jmir.3863>.
94. Rekatsinas, T.; Ghosh, S.; Mekaru, S.R.; Nsoesie, E.O.; Brownstein, J.S.; Getoor, L.; Ramakrishnan, N. SourceSeer: Forecasting Rare Disease Outbreaks Using Multiple Data Sources. In Proceedings of the Proceedings of the 2015 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2015, pp. 379–387. <https://doi.org/10.1137/1.9781611974010.43>.
95. Rekatsinas, T.; Ghosh, S.; Mekaru, S.R.; Nsoesie, E.O.; Brownstein, J.S.; Getoor, L.; Ramakrishnan, N. Forecasting rare disease outbreaks from open source indicators. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **2017**, *10*, 136–150. <https://doi.org/10.1002/sam.11337>.
96. Thapen, N.; Simmie, D.; Hankin, C.; Gillard, J. DEFENDER: Detecting and Forecasting Epidemics Using Novel Data-Analytics for Enhanced Response. *PLOS ONE* **2016**, *11*, e0155417. <https://doi.org/10.1371/journal.pone.0155417>.
97. Ghosh, S.; Chakraborty, P.; Cohn, E.; Brownstein, J.S.; Ramakrishnan, N. Characterizing diseases from unstructured text: A vocabulary driven word2vec approach. In Proceedings of the Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, 2016, pp. 1129–1138.
98. Kim, E.H.J.; Jeong, Y.K.; Kim, Y.; Kang, K.Y.; Song, M. Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science* **2016**, *42*, 763–781. <https://doi.org/10.1177/0165551515608733>.
99. Ghosh, S. News Analytics for Global Infectious Disease Surveillance. PhD thesis, Virginia Tech, 2017.

100. Ghosh, S.; Chakraborty, P.; Nsoesie, E.O.; Cohn, E.; Mekaru, S.R.; Brownstein, J.S.; Ramakrishnan, N. Temporal Topic Modeling to Assess Associations between News Trends and Infectious Disease Outbreaks. *Scientific Reports* **2017**, *7*, 40841. <https://doi.org/10.1038/srep40841>.
101. Villanes, A.; Griffiths, E.; Rappa, M.; Healey, C.G. Dengue Fever Surveillance in India Using Text Mining in Public Media. *The American Journal of Tropical Medicine and Hygiene* **2018**, *98*, 181–191. <https://doi.org/10.4269/ajtmh.17-0253>.
102. Huang, M.; ElTayeb, O.; Zolnoori, M.; Yao, L. Public Opinions Toward Diseases: Infodemiological Study on News Media Data. *Journal of Medical Internet Research* **2018**, *20*, e10047. <https://doi.org/10.2196/10047>.
103. Bouzembrak, Y.; Steen, B.; Neslo, R.; Linge, J.; Mojtahed, V.; Marvin, H. Development of food fraud media monitoring system based on text mining. *Food Control* **2018**, *93*, 283–296. <https://doi.org/10.1016/j.foodcont.2018.06.003>.
104. Balashankar, A.; Dugar, A.; Subramanian, L.; Fraiberger, S. Reconstructing the MERS disease outbreak from news. In Proceedings of the Proceedings of the Conference on Computing & Sustainable Societies - COMPASS 19, Accra, Ghana, 2019; pp. 272–280. <https://doi.org/10.1145/3314344.3332498>.
105. Cho, K.W.; Kim, S.Y.; Woo, Y.W. Analysis of Women's Health Online News Articles Using Topic Modeling. *Osong Public Health and Research Perspectives* **2019**, *10*, 158–169. <https://doi.org/10.24171/j.phrp.2019.10.3.07>.
106. Kim, M.G.; Huh, S.; Han, N.; Kim, J.H.; Kim, K.; Lee, E.; Kim, I.W.; Oh, J.M. Analysis of Issues and Future Trends Impacting Drug Safety in South Korea. *International Journal of Environmental Research and Public Health* **2019**, *16*, 3368. <https://doi.org/10.3390/ijerph16183368>.
107. Huang, J.Y.; Lee, H.M. Knowledge Discovery Model in Chinese Industrial News. In Proceedings of the Proceedings of the Second International Conference on Electronic Business (ICEB-2002), 2002, pp. 412–414.
108. Hong, T.; Han, I. Knowledge-based data mining of news information on the Internet using cognitive maps and neural networks. *Expert Systems with Applications* **2002**, *23*, 1–8. [https://doi.org/10.1016/S0957-4174\(02\)00022-2](https://doi.org/10.1016/S0957-4174(02)00022-2).
109. Wei, C.P.; Lee, Y.H. Event detection from online news documents for supporting environmental scanning. *Decision Support Systems* **2004**, *36*, 385–401. [https://doi.org/10.1016/S0167-9236\(03\)00028-9](https://doi.org/10.1016/S0167-9236(03)00028-9).
110. Zhang, D.; Simoff, S.J. Informing the Curious Negotiator: Automatic News Extraction from the Internet. In *Data Mining*; Hutchison, D.; Kanade, T.; Kittler, J.; Kleinberg, J.M.; Mattern, F.; Mitchell, J.C.; Naor, M.; Nierstrasz, O.; Pandu Rangan, C.; Steffen, B.; et al., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; Vol. 3755, pp. 176–191. [https://doi.org/10.1007/11677437\\_14](https://doi.org/10.1007/11677437_14).
111. Yu, W.B.; Lea, B.R.; Guruswamy, B. A Theoretic Framework Integrating Text Mining and Energy Demand Forecasting. *IJEBM* **2007**, *5*, 211–224.
112. Yeh, P.Z.; Kass, A. Capturing the Semantics of Online News Sources for Business Intelligence Applications. In Proceedings of the 2008 20th IEEE International Conference on Tools with Artificial Intelligence, Dayton, OH, USA, 2008; pp. 111–117. <https://doi.org/10.1109/ICTAI.2008.65>.
113. Liu, D.R.; Shih, M.J.; Liao, C.J.; Lai, C.H. Mining the change of event trends for decision support in environmental scanning. *Expert Systems with Applications* **2009**, *36*, 972–984. <https://doi.org/10.1016/j.eswa.2007.10.016>.
114. Dey, L.; Haque, S.M.; Khurdiya, A.; Shroff, G. Acquiring competitive intelligence from social media. In Proceedings of the Proceedings of the 2011 Joint Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data - MOCR\_AND '11, Beijing, China, 2011; p. 1. <https://doi.org/10.1145/2034617.2034621>.
115. Wex, F.; Widder, N.; Liebmann, M.; Neumann, D. Early Warning of Impending Oil Crises Using the Predictive Power of Online News Stories. In Proceedings of the 2013 46th Hawaii International Conference on System Sciences, Wailea, HI, USA, 2013; pp. 1512–1521. <https://doi.org/10.1109/HICSS.2013.186>.
116. Özyirmidokuz, E.K. Mining Unstructured Turkish Economy News Articles. *Procedia Economics and Finance* **2014**, *16*, 320–328. [https://doi.org/10.1016/S2212-5671\(14\)00809-0](https://doi.org/10.1016/S2212-5671(14)00809-0).
117. Mao, H.; Wang, K.; Ma, R.; Gao, Y.; Li, Y.; Chen, K.; Xie, D.; Zhu, W.; Wang, T.; Wang, H. An Automatic News Analysis and Opinion Sharing System for Exchange Rate Analysis. In Proceedings of the 2014 IEEE 11th International Conference on e-Business Engineering, Guangzhou, China, 2014; pp. 303–307. <https://doi.org/10.1109/ICEBE.2014.59>.
118. Abdelkader, A.; Hand, E.; Samet, H. Brands in NewsStand: spatio-temporal browsing of business news. In Proceedings of the Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '15, Bellevue, Washington, 2015; pp. 1–4. <https://doi.org/10.1145/2820783.2820795>.

119. Verma, I.; Dey, L.; Srinivasan, R.S.; Singh, L. Event Detection from Business News. In *Pattern Recognition and Machine Intelligence*; Kryszkiewicz, M.; Bandyopadhyay, S.; Rybinski, H.; Pal, S.K., Eds.; Springer International Publishing: Cham, 2015; Vol. 9124, pp. 575–585. Series Title: Lecture Notes in Computer Science, [https://doi.org/10.1007/978-3-319-19941-2\\_55](https://doi.org/10.1007/978-3-319-19941-2_55).
120. Dasgupta, T.; Dey, L.; Dey, P.; Saha, R. A framework for mining enterprise risk and risk factors from news documents. In Proceedings of the Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations, 2016, pp. 180–184.
121. Chan, Y.Y.; Qu, H. FinaVistory: Using Narrative Visualization to explain social and Economic relationships in financial news. In Proceedings of the 2016 International Conference on Big Data and Smart Computing (BigComp), Hong Kong, China, 2016; pp. 32–39. <https://doi.org/10.1109/BIGCOMP.2016.7425798>.
122. Rönnqvist, S.; Sarlin, P. Bank distress in the news: Describing events through deep learning. *Neurocomputing* **2017**, *264*, 57–70.
123. Chang, T.M.; Hu, G.H.; Hsu, M.F.; Lin, K.P. Integration of Social Media News Mining and Text Mining Techniques to Determine a Corporate's Competitive Edge. In Proceedings of the Pacific Asia Conference on Information Systems (PACIS). Association For Information Systems, 2017.
124. Yamamoto, A.; Miyamura, Y.; Nakata, K.; Okamoto, M. Company Relation Extraction from Web News Articles for Analyzing Industry Structure. In Proceedings of the 2017 IEEE 11th International Conference on Semantic Computing (ICSC), San Diego, CA, USA, 2017; pp. 89–92. <https://doi.org/10.1109/ICSC.2017.25>.
125. Benetka, J.R.; Balog, K.; Nørnvåg, K. Towards Building a Knowledge Base of Monetary Transactions from a News Collection. *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL) 2017*, pp. 1–10. arXiv: 1709.05743, <https://doi.org/10.1109/JCDL.2017.7991575>.
126. Moon, H.; Kim, S.; Yang, T.; Kim, K. Big Data Analysis on Gambling Related News in South Korea. In Proceedings of the 2017 IEEE International Congress on Big Data (BigData Congress), Honolulu, HI, USA, 2017; pp. 541–544. <https://doi.org/10.1109/BigDataCongress.2017.82>.
127. Kim, D.; Kim, S. Sustainable Supply Chain Based on News Articles and Sustainability Reports: Text Mining with Leximancer and DICTION. *Sustainability* **2017**, *9*, 1008. <https://doi.org/10.3390/su9061008>.
128. Fu, Y.; Hao, J.X.; (Robert) Li, X.; Hsu, C.H. Predictive Accuracy of Sentiment Analytics for Tourism: A Metalearning Perspective on Chinese Travel News. *Journal of Travel Research* **2019**, *58*, 666–679. <https://doi.org/10.1177/0047287518772361>.
129. Agarwal, A.; Gupta, A.; Kumar, A.; Tamilselvam, S.G. Learning risk culture of banks using news analytics. *European Journal of Operational Research* **2019**, *277*, 770–783. <https://doi.org/10.1016/j.ejor.2019.02.045>.
130. Kanhabua, N.; Blanco, R.; Matthews, M. Ranking related news predictions. In Proceedings of the Proceedings of the 34th international ACM SIGIR conference on Research and development in Information - SIGIR '11, Beijing, China, 2011; p. 755. <https://doi.org/10.1145/2009916.2010018>.
131. Ho, S.S.; Lieberman, M.; Wang, P.; Samet, H. Mining future spatiotemporal events and their sentiment from online news articles for location-aware recommendation system. In Proceedings of the Proceedings of the First ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems - MobiGIS '12, Redondo Beach, California, 2012; p. 25. <https://doi.org/10.1145/2442810.2442816>.
132. Radinsky, K.; Davidovich, S.; Markovitch, S. Learning causality for news events prediction. In Proceedings of the Proceedings of the 21st international conference on World Wide Web - WWW '12, Lyon, France, 2012; pp. 909–918. <https://doi.org/10.1145/2187836.2187958>.
133. Radinsky, K.; Horvitz, E. Mining the web to predict future events. In Proceedings of the Proceedings of the sixth ACM international conference on Web search and data mining - WSDM '13, Rome, Italy, 2013; p. 255. <https://doi.org/10.1145/2433396.2433431>.
134. Yeon, H.; Jang, Y. Predictive Visual Analytics using Topic Composition. In Proceedings of the Proceedings of the 8th International Symposium on Visual Information Communication and Interaction - VINCI '15, Tokyo, AA, Japan, 2015; pp. 1–8. <https://doi.org/10.1145/2801040.2801054>.
135. Kayser, V.; Blind, K. Extending the knowledge base of foresight: The contribution of text mining. *Technological Forecasting and Social Change* **2017**, *116*, 208–215. <https://doi.org/10.1016/j.techfore.2016.10.017>.
136. Dami, S.; Barforoush, A.A.; Shirazi, H. News events prediction using Markov logic networks. *Journal of Information Science* **2018**, *44*, 91–109. <https://doi.org/10.1177/0165551516673285>.
137. Yarrabelly, N.; Karlapalem, K. Estimating Credibility of News Authors from their WIKI Validated Predictions. *NewsIR@ ECIR* **2018**, *2079*, 12–17.

138. Au Yeung, C.m.; Jatowt, A. Studying How the Past is Remembered: Towards Computational History through Large Scale Text Mining. In Proceedings of the Proceedings of the 20th ACM international conference on Information and knowledge management, 2011, pp. 1231–1240.
139. Thompson, P.; Nawaz, R.; Korkontzelos, I.; Black, W.; McNaught, J.; Ananiadou, S. News search using discourse analytics. In Proceedings of the 2013 Digital Heritage International Congress (DigitalHeritage), Marseille, France, 2013; pp. 597–604. <https://doi.org/10.1109/DigitalHeritage.2013.6743801>.
140. Yzaguirre, A.; Smit, M.; Warren, R. Newspaper archives + text mining = rich sources of historical geo-spatial data. *IOP Conference Series: Earth and Environmental Science* **2016**, *34*, 012043. <https://doi.org/10.1088/1755-1315/34/1/012043>.
141. Shook, E.; Leetaru, K.; Cao, G.; Padmanabhan, A.; Wang, S. Happy or not: Generating topic-based emotional heatmaps for Culturomics using CyberGIS. In Proceedings of the 2012 IEEE 8th International Conference on E-Science, Chicago, IL, USA, 2012; pp. 1–6. <https://doi.org/10.1109/eScience.2012.6404440>.
142. Piskorski, J.; Tanev, H.; Oezden Wennerberg, P. Extracting Violent Events From On-Line News for Ontology Population. In *Business Information Systems*; Abramowicz, W., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; Vol. 4439, pp. 287–300. [https://doi.org/10.1007/978-3-540-72035-5\\_22](https://doi.org/10.1007/978-3-540-72035-5_22).
143. Yonamine, J.E. Predicting Future Levels of Violence in Afghanistan Districts Using GDELT. *Unpublished Manuscript* **2013**.
144. Csala, D. *Insurgent Dynamics: A systematic analysis of social unrest*; 2014; p. 22.
145. Holliday, M.; Holden, C. Advanced Web-Based Temporal Analytics for Arms Control. *AAAS Science & Diplomacy Journal*.
146. Yang, C.C.; Shi, X.; Wei, C.P. Tracing the Event Evolution of Terror Attacks from On-Line News. In *Intelligence and Security Informatics*; Hutchison, D.; Kanade, T.; Kittler, J.; Kleinberg, J.M.; Mattern, F.; Mitchell, J.C.; Naor, M.; Nierstrasz, O.; Pandu Rangan, C.; Steffen, B.; et al., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; Vol. 3975, pp. 343–354. [https://doi.org/10.1007/11760146\\_30](https://doi.org/10.1007/11760146_30).
147. Ahmed, S.T.; Bhindwale, R.; Davulcu, H. Tracking terrorism news threads by extracting event signatures. In Proceedings of the 2009 IEEE International Conference on Intelligence and Security Informatics, Richardson, TX, USA, 2009; pp. 182–184. <https://doi.org/10.1109/ISI.2009.5137296>.
148. Wiil, U.K., Ed. *Counterterrorism and open source intelligence*; Number v. 2 in Lecture notes in social networks, Springer: Wien ; New York, 2011. OCLC: ocn690089099.
149. Nizamani, S.; Memon, N. Analyzing News Summaries for Identification of Terrorism Incident. *Educ Res Int* **2014**, *3*.
150. Toure, I.; Gangopadhyay, A. Analyzing real time terrorism data. In Proceedings of the 2015 IEEE International Symposium on Technologies for Homeland Security (HST), Waltham, MA, USA, 2015; pp. 1–4. <https://doi.org/10.1109/THS.2015.7225299>.
151. Sheremetyeva, S.; Zinoveva, A. Ontological Analysis of E-News: A Case for Terrorism Domain. In Proceedings of the Proceedings of the IS-2019 Conference; Sosnin, P.; Maklaev, V.; Sosnina, E., Eds., Ulyanovsk, Russia, 2019; p. 12.
152. Piskorski, J.; Tanev, H.; Atkinson, M.; van der Goot, E.; Zavarella, V. Online News Event Extraction for Global Crisis Surveillance. In *Transactions on Computational Collective Intelligence V*; Nguyen, N.T., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; Vol. 6910, pp. 182–212. [https://doi.org/10.1007/978-3-642-24016-4\\_10](https://doi.org/10.1007/978-3-642-24016-4_10).
153. Kwak, H.; An, J. A First Look at Global News Coverage of Disasters by Using the GDELT Dataset. In *Social Informatics*; Aiello, L.M.; McFarland, D., Eds.; Springer International Publishing: Cham, 2014; Vol. 8851, pp. 300–308. [https://doi.org/10.1007/978-3-319-13734-6\\_22](https://doi.org/10.1007/978-3-319-13734-6_22).
154. Kwak, H.; An, J. Understanding news geography and major determinants of global news coverage of disasters. *arXiv preprint arXiv:1410.3710* **2014**.
155. Racette, M.P.; Smith, C.T.; Cunningham, M.P.; Heekin, T.A.; Lemley, J.P.; Mathieu, R.S. Improving situational awareness for humanitarian logistics through predictive modeling. In Proceedings of the 2014 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA, 2014; pp. 334–339. <https://doi.org/10.1109/SIEDS.2014.6829918>.
156. Kedzie, C.; McKeown, K.; Diaz, F. Predicting Salient Updates for Disaster Summarization. In Proceedings of the Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 2015; pp. 1608–1617. <https://doi.org/10.3115/v1/P15-1155>.

157. Teodorescu, H.N. Using Analytics and Social Media for Monitoring and Mitigation of Social Disasters. *Procedia Engineering* **2015**, *107*, 325–334. <https://doi.org/10.1016/j.proeng.2015.06.088>.
158. Yzaguirre, A.; Warren, R.; Smit, M. Detecting environmental disasters in digital news archives. In Proceedings of the 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 2015; pp. 2027–2035. <https://doi.org/10.1109/BigData.2015.7363984>.
159. Karuna, P.; Rana, M.; Purohit, H. CitizenHelper: A Streaming Analytics System to Mine Citizen and Web Data for Humanitarian Organizations. In Proceedings of the ICWSM, 2017, pp. 729–730.
160. Pandey, R.; Purohit, H. CitizenHelper-Adaptive: Expert-Augmented Streaming Analytics System for Emergency Services and Humanitarian Organizations. In Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, 2018; pp. 630–633. <https://doi.org/10.1109/ASONAM.2018.8508374>.
161. Nourbakhsh, A.; Li, Q.; Liu, X.; Shah, S. "Breaking" Disasters: Predicting and Characterizing the Global News Value of Natural and Man-made Disasters. *arXiv:1709.02510 [physics]* **2017**. arXiv: 1709.02510.
162. Yulianto, R.; Mariyah, S. Building automatic mind map generator for natural disaster news in Bahasa Indonesia. In Proceedings of the 2017 International Conference on Information Technology Systems and Innovation (ICITSI), Bandung, Indonesia, 2017; pp. 177–182. <https://doi.org/10.1109/ICITSI.2017.8267939>.
163. Liu, X.; Guo, H.; Lin, Y.; Li, Y.; Hou, J. Analyzing Spatial-Temporal Distribution of Natural Hazards in China by Mining News Sources. *Natural Hazards Review* **2018**, *19*, 04018006. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000291](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000291).
164. Banujan, K.; Kumara, B.T.; Paik, I. Twitter and Online News analytics for Enhancing Post-Natural Disaster Management Activities. In Proceedings of the 2018 9th International Conference on Awareness Science and Technology (iCAST), Fukuoka, 2018; pp. 302–307. <https://doi.org/10.1109/ICAwST.2018.8517195>.
165. Trumbo, C. Constructing climate change: claims and frames in US news coverage of an environmental issue. *Public Understanding of Science* **1996**, *5*, 269–283. <https://doi.org/10.1088/0963-6625/5/3/006>.
166. Rivera, S.J.; Minsker, B.S.; Work, D.B.; Roth, D. A text mining framework for advancing sustainability indicators. *Environmental Modelling & Software* **2014**, *62*, 128–138. <https://doi.org/10.1016/j.envsoft.2014.08.016>.
167. Hyun, Y.; Kim, J.S.; Jeong, J.W.; Yun, S.; Lee, M.S. Text mining on internet-news regarding climate change and food. *Journal of the Korean Data and Information Science Society* **2015**, *26*, 419–427. <https://doi.org/10.7465/jkdi.2015.26.2.419>.
168. Hori, S. An exploratory analysis of the text mining of news articles about "water and society". In *WIT Transactions on The Built Environment*, 1 ed.; Brebbia, C.A., Ed.; WIT Press, 2015; Vol. 1, pp. 501–508. <https://doi.org/10.2495/SD150441>.
169. Albrieu, R.; Palazzo, G. Mapping social conflicts in natural resources. A text-mining study in mining activities. MPRA Paper 93155, University Library of Munich, Germany, 2016.
170. Boussalis, C.; Coan, T.G. Text-mining the signals of climate change doubt. *Global Environmental Change* **2016**, *36*, 89–100. <https://doi.org/10.1016/j.gloenvcha.2015.12.001>.
171. Conley-Zilkic, B. Preventing and Responding to Mass Atrocities. *African Politics* **2016**, p. 7.
172. Roby, N.A.; Gonzales, P.; Quesnel, K.J.; Ajami, N.K. A novel search algorithm for quantifying news media coverage as a measure of environmental issue salience. *Environmental Modelling & Software* **2018**, *101*, 249–255. <https://doi.org/10.1016/j.envsoft.2017.12.012>.
173. Gulli, A. The anatomy of a news search engine. In Proceedings of the Special interest tracks and posters of the 14th international conference on World Wide Web - WWW '05, Chiba, Japan, 2005; p. 880. <https://doi.org/10.1145/1062745.1062778>.
174. Lloyd, L.; Kechagias, D.; Skiena, S. Lydia: A System for Large-Scale News Analysis. In *String Processing and Information Retrieval*; Hutchison, D.; Kanade, T.; Kittler, J.; Kleinberg, J.M.; Mattern, F.; Mitchell, J.C.; Naor, M.; Nierstrasz, O.; Pandu Rangan, C.; Steffen, B.; et al., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; Vol. 3772, pp. 161–166. [https://doi.org/10.1007/11575832\\_18](https://doi.org/10.1007/11575832_18).
175. Mehler, A.; Bao, Y.; Li, X.; Wang, Y.; Skiena, S. Spatial Analysis of News Sources. *IEEE Transactions on Visualization and Computer Graphics* **2006**, *12*, 765–772. <https://doi.org/10.1109/TVCG.2006.179>.
176. Choudhary, R.; Mehta, S.; Bagchi, A.; Balakrishnan, R. Towards characterization of actor evolution and interactions in news corpora. In Proceedings of the European Conference on Information Retrieval. Springer, 2008, pp. 422–429.
177. Picchi, E.; Cucurullo, S.; Sassolini, E.; Bertagna, F. Mining the news with semantic press. *Proceedings of LangTech* **2008**, pp. 141–144.

178. Teitler, B.E.; Lieberman, M.D.; Panozzo, D.; Sankaranarayanan, J.; Samet, H.; Sperling, J. NewsStand: a new view on news. In Proceedings of the Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems - GIS '08, Irvine, California, 2008; p. 1. <https://doi.org/10.1145/1463434.1463458>.
179. Wanner, F.; Rohrdantz, C.; Mansmann, F.; Stoffel, A.; Oelke, D.; Krstajic, M.; Keim, D.; Luo, D.; Yang, J.; Atkinson, M. Large-scale comparative sentiment analysis of news articles. In Proceedings of the InfoVis, 2009.
180. Daniel, A.; KRSTAJIC, M.; Peter, B.; Oelke, D.; Mansmann, F. Methods for interactive exploration of large-scale news streams. *Web intelligence and security: advances in data and text mining techniques for detecting and preventing terrorist activities on the web* **2010**, *27*, 139.
181. Krstajić, M.; Bertini, E.; Mansmann, F.; Keim, D.A. Visual analysis of news streams with article threads. In Proceedings of the Proceedings of the First International Workshop on Novel Data Stream Pattern Mining Techniques - StreamKDD '10, Washington, D.C., 2010; pp. 39–46. <https://doi.org/10.1145/1833280.1833286>.
182. Krstajić, M.; Najm-Araghi, M.; Mansmann, F.; Keim, D.A. Incremental visual text analytics of news story development. In Proceedings of the Visualization and Data Analysis 2012; Wong, P.C.; Kao, D.L.; Hao, M.C.; Chen, C.; Kosara, R.; Livingston, M.A.; Park, J.; Roberts, I., Eds. International Society for Optics and Photonics, 2012, Vol. 8294, p. 829407. <https://doi.org/10.1117/12.912456>.
183. Krstajić, M.; Najm-Araghi, M.; Mansmann, F.; Keim, D.A. Story Tracker: Incremental visual text analytics of news story development. *Information Visualization* **2013**, *12*, 308–323. <https://doi.org/10.1177/1473871613493996>.
184. Glavaš, G.; Štajner, S. Event-centered simplification of news stories. In Proceedings of the Proceedings of the Student Research Workshop associated with RANLP 2013, 2013, pp. 71–78.
185. Gao, T.; Hullman, J.R.; Adar, E.; Hecht, B.; Diakopoulos, N. NewsViews: an automated pipeline for creating custom geovisualizations for news. In Proceedings of the Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14, Toronto, Ontario, Canada, 2014; pp. 3005–3014. <https://doi.org/10.1145/2556288.2557228>.
186. Hoffart, J.; Milchevski, D.; Weikum, G. AESTHETICS: Analytics with Strings, Things, and Cats. In Proceedings of the Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management - CIKM '14, Shanghai, China, 2014; pp. 2018–2020. <https://doi.org/10.1145/2661829.2661835>.
187. Tanisha, L.F.; Pathik, B.B.; Khan, M.H.; Habib, M.M. Analyzing and visualizing news trends over time. In Proceedings of the 2014 IEEE International Conference on Industrial Engineering and Engineering Management, Selangor Darul Ehsan, Malaysia, 2014; pp. 307–311. <https://doi.org/10.1109/IEEM.2014.7058649>.
188. Ploch, D. Intelligent News Aggregator for German with Sentiment Analysis. In *Smart Information Systems*; Hopfgartner, F., Ed.; Springer International Publishing: Cham, 2015; pp. 5–46. [https://doi.org/10.1007/978-3-319-14178-7\\_1](https://doi.org/10.1007/978-3-319-14178-7_1).
189. Steinberger, J. MediaGist: A Cross-lingual Analyser of Aggregated News and Commentaries. In Proceedings of the Proceedings of ACL-2016 System Demonstrations, Berlin, Germany, 2016; pp. 145–150. <https://doi.org/10.18653/v1/P16-4025>.
190. Saleiro, P.; Teixeira, J.; Soares, C.; Oliveira, E. TimeMachine: Entity-centric Search and Visualization of News Archives. In Proceedings of the European Conference on Information Retrieval. Springer, 2016, pp. 845–848.
191. Liu, S.; Yin, J.; Wang, X.; Cui, W.; Cao, K.; Pei, J. Online Visual Analytics of Text Streams. *IEEE Transactions on Visualization and Computer Graphics* **2016**, *22*, 2451–2466. arXiv: 1512.04042, <https://doi.org/10.1109/TVCG.2015.2509990>.
192. Hamborg, F.; Meuschke, N.; Breiting, C.; Gipp, B. Identification and Analysis of Media Bias in News Articles - The Impact of Objectivity and Believability on Corporate Decision Making and Performance. In Proceedings of the ISI, 2017.
193. Hamborg, F.; Meuschke, N.; Gipp, B. Bias-aware news analysis using matrix-based news aggregation. *International Journal on Digital Libraries* **2018**. <https://doi.org/10.1007/s00799-018-0239-9>.
194. Laban, P.; Hearst, M. newsLens: building and visualizing long-ranging news stories. In Proceedings of the Proceedings of the Events and Stories in the News Workshop, Vancouver, Canada, 2017; pp. 1–9. <https://doi.org/10.18653/v1/W17-2701>.

195. Hu, Y.; Ye, X.; Shaw, S.L. Extracting and analyzing semantic relatedness between cities using news articles. *International Journal of Geographical Information Science* **2017**, *31*, 2427–2451. <https://doi.org/10.1080/13658816.2017.1367797>.
196. Schubert, E.; Spitz, A.; Gertz, M. Exploring Significant Interactions in Live News. *NewsIR@ ECIR* **2018**, 2079, 39–44.
197. Poghosyan, G.; Ifrim, G. SocialTree: Socially Augmented Structured Summaries of News Stories. In Proceedings of the Proceedings of the 30th ACM Conference on Hypertext and Social Media - HT '19, Hof, Germany, 2019; pp. 153–162. <https://doi.org/10.1145/3342220.3343668>.
198. Lim, C.; Maglio, P.P. Data-Driven Understanding of Smart Service Systems Through Text Mining. *Service Science* **2018**, *10*, 154–180. <https://doi.org/10.1287/serv.2018.0208>.
199. Mejia, C.; Kajikawa, Y. The field of social robotics as means of technology selection to address country specific social issues. In Proceedings of the 2016 Portland International Conference on Management of Engineering and Technology (PICMET), Honolulu, HI, USA, 2016; pp. 2913–2921. <https://doi.org/10.1109/PICMET.2016.7806653>.
200. Yoshioka, M. Analyzing Multiple News Sites by Contrasting Articles. In Proceedings of the 2008 IEEE International Conference on Signal Image Technology and Internet Based Systems, Bali, Indonesia, 2008; pp. 45–51. <https://doi.org/10.1109/SITIS.2008.42>.
201. Diakopoulos, N.; Zhang, A.X.; Salway, A. Visual analytics of media frames in online news and blogs. In Proceedings of the Proc. IEEE InfoVis Workshop on Text Visualization, 2013.
202. Hensinger, E.; Flaounas, I.; Cristianini, N. Modelling and predicting news popularity. *Pattern Analysis and Applications* **2013**, *16*, 623–635. <https://doi.org/10.1007/s10044-012-0314-6>.
203. Toraman, C.; Can, F. A front-page news-selection algorithm based on topic modelling using raw text. *Journal of Information Science* **2015**, *41*, 676–685. <https://doi.org/10.1177/0165551515589069>.
204. Liu, Q.; Zhou, M.; Zhao, X. Understanding News 2.0: A framework for explaining the number of comments from readers on online news. *Information & Management* **2015**, *52*, 764–776. <https://doi.org/10.1016/j.im.2015.01.002>.
205. Cheeks, L.H.; Stepien, T.L.; Wald, D.M.; Gaffar, A. Discovering News Frames: An Approach for Exploring Text, Content, and Concepts in Online News Sources. *International Journal of Multimedia Data Engineering and Management* **2016**, *7*, 45–62. <https://doi.org/10.4018/IJMDEM.2016100103>.
206. Sohrabi, B.; Vanani, I.R.; Namavar, M. Investigation of Trends and Analysis of Hidden New Patterns in Prominent News Agencies of Iran Using Data Mining and Text Mining Algorithms **2019**. p. 24.
207. Lee, C. How are 'immigrant workers' represented in Korean news reporting?-A text mining approach to critical discourse analysis. *Digital Scholarship in the Humanities* **2019**, *34*, 82–99. <https://doi.org/10.1093/llc/fqy017>.
208. Ye, J.; Skiena, S. MediaRank: Computational ranking of online news sources. In Proceedings of the Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 2469–2477.
209. Liao, Y.; Wang, S.; Han, E.H.S.; Lee, J.; Lee, D. Characterization and Early Detection of Evergreen News Articles. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, 2019, pp. 552–568.
210. Yoon, I.J.; Han, K.D.; Kim, H. Hate Speech against Immigrants in Korea. *Asia Review* **2018**, *8*, 259–288. <https://doi.org/10.24987/SNUACAR.2018.08.8.1.259>.
211. Ali, O.; Flaounas, I.; De Bie, T.; Mosdell, N.; Lewis, J.; Cristianini, N. Automating news content analysis: an application to gender bias and readability. In Proceedings of the JMLR: Workshop and Conference Proceedings; Diethe, T.; Cristianini, N.; Shawe-Taylor, J., Eds., 2010, Vol. 11, pp. 36–43.
212. Card, D.; Gross, J.; Boydston, A.; Smith, N.A. Analyzing Framing through the Casts of Characters in the News. In Proceedings of the Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas, 2016; pp. 1410–1420. <https://doi.org/10.18653/v1/D16-1148>.
213. Agrawal, A.; Sahdev, R.; Davoudi, H.; Khonsari, F.; An, A.; McGrath, S. Detecting the Magnitude of Events from News Articles. In Proceedings of the 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Omaha, NE, USA, 2016; pp. 177–184. <https://doi.org/10.1109/WI.2016.0034>.
214. Khonsari, F. Mining Large-scale News Articles for Predicting Forced Migration via Machine Learning Techniques. PhD thesis, York University, 2018.
215. Singh, L.; Wahedi, L.; Wang, Y.; Wei, Y.; Kirov, C.; Martin, S.; Donato, K.; Liu, Y.; Kawintiranon, K. Blending Noisy Social Media Signals with Traditional Movement Variables to Predict Forced Migration. In Proceedings



- of the Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '19, Anchorage, AK, USA, 2019; pp. 1975–1983. <https://doi.org/10.1145/3292500.3330774>.
216. Fister, I.; Fister, D.; Rauter, S.; Mlakar, U.; Brest, J.; Fister, I. Deep Analytics Based on Triathlon Athletes' Blogs and News. In *Recent Advances in Soft Computing*; Matoušek, R., Ed.; Springer International Publishing: Cham, 2019; Vol. 837, pp. 279–289. [https://doi.org/10.1007/978-3-319-97888-8\\_25](https://doi.org/10.1007/978-3-319-97888-8_25).
217. Montes, M.; Gelbukh, A.; López, A.L.; et al. Mining the news: trends, associations, and deviations. *Computación y Sistemas* **2001**, *5*, 14–24.
218. Leetaru, K. Culturomics 2.0: Forecasting large-scale human behavior using global news media tone in time and space. *First Monday* **2011**.
219. Phua, C.; Feng, Y.; Ji, J.; Soh, T. Visual and Predictive Analytics on Singapore News: Experiments on GDELT, Wikipedia, and ŠTI. *arXiv e-prints* **2014**, p. arXiv:1404.1996, [arXiv:cs.OH/1404.1996].
220. Sheng Bi.; Gao, J.; Yandon Wang.; Yinhe Cao. A contrast of the degree of activity among the three major powers, USA, China, and Russia: Insights from media reports. In Proceedings of the 2015 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC), Nanjing, China, 2015; pp. 38–42. <https://doi.org/10.1109/BESC.2015.7365955>.
221. Yoshioka, M.; Kando, N. Comparative Analysis of GDELT Data Using the News Site Contrast System. In Proceedings of the NewsIR@ ECIR, 2016, pp. 63–65.
222. Salloum, S.A.; Al-Emran, M.; Abdallah, S.; Shaalan, K. Analyzing the Arab Gulf Newspapers Using Text Mining Techniques. In *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2017*; Hassanien, A.E.; Shaalan, K.; Gaber, T.; Tolba, M.F., Eds.; Springer International Publishing: Cham, 2018; Vol. 639, pp. 396–405. [https://doi.org/10.1007/978-3-319-64861-3\\_37](https://doi.org/10.1007/978-3-319-64861-3_37).
223. Suh, J.H. SocialTERM-Extractor: Identifying and Predicting Social-Problem-Specific Key Noun Terms from a Large Number of Online News Articles Using Text Mining and Machine Learning Techniques. *Sustainability* **2019**, *11*, 196. <https://doi.org/10.3390/su11010196>.
224. Shinyama, Y.; Sekine, S. Named entity discovery using comparable news articles. In Proceedings of the Proceedings of the 20th international conference on Computational Linguistics - COLING '04, Geneva, Switzerland, 2004; pp. 848–es. <https://doi.org/10.3115/1220355.1220477>.
225. Newman, D.; Chemudugunta, C.; Smyth, P.; Steyvers, M. Analyzing Entities and Topics in News Articles Using Statistical Topic Models. In *Intelligence and Security Informatics*; Hutchison, D.; Kanade, T.; Kittler, J.; Kleinberg, J.M.; Mattern, F.; Mitchell, J.C.; Naor, M.; Nierstrasz, O.; Pandu Rangan, C.; Steffen, B.; et al., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; Vol. 3975, pp. 93–104. [https://doi.org/10.1007/11760146\\_9](https://doi.org/10.1007/11760146_9).
226. Pouliquen, B.; Steinberger, R.; Ignat, C.; Oellinger, T. Building and Displaying Name Relations using Automatic Unsupervised Analysis of Newspaper Articles. *arXiv e-prints* **2006**, p. cs/0609066, [arXiv:cs.CL/cs/0609066].
227. Liu, M.; Liu, Y.; Xiang, L.; Chen, X.; Yang, Q. Extracting Key Entities and Significant Events from Online Daily News. In *Intelligent Data Engineering and Automated Learning - IDEAL 2008*; Fyfe, C.; Kim, D.; Lee, S.Y.; Yin, H., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2008; Vol. 5326, pp. 201–209. [https://doi.org/10.1007/978-3-540-88906-9\\_26](https://doi.org/10.1007/978-3-540-88906-9_26).
228. Atkinson, M.; Van der Goot, E. Near real time information mining in multilingual news. In Proceedings of the Proceedings of the 18th international conference on World wide web - WWW '09, Madrid, Spain, 2009; p. 1153. <https://doi.org/10.1145/1526709.1526903>.
229. Yang, C.; Xiaodong Shi.; Chih-Ping Wei. Discovering Event Evolution Graphs From News Corpora. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* **2009**, *39*, 850–863. <https://doi.org/10.1109/TSMCA.2009.2015885>.
230. Tran, M.V.; Nguyen, M.H.; Nguyen, S.Q.; Nguyen, M.T.; Phan, X.H. VnLoc: A Real – Time News Event Extraction Framework for Vietnamese. In Proceedings of the 2012 Fourth International Conference on Knowledge and Systems Engineering, Danang, Vietnam, 2012; pp. 161–166. <https://doi.org/10.1109/KSE.2012.34>.
231. Park, S.; Lee, K.S.; Song, J. Contrasting Opposing Views of News Articles on Contentious Issues. In Proceedings of the Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies, 2011, pp. 340–349.
232. Lin, Y.F.; Kao, H.Y. The Retrieval of Important News Stories by Influence Propagation among Communities and Categories. In Proceedings of the 2012 IEEE/WIC/ACM International Conferences on Web Intelligence

- and Intelligent Agent Technology, Macau, China, 2012; pp. 32–39. <https://doi.org/10.1109/WI-IAT.2012.236>.
233. Nakashole, N.; Tylenda, T.; Weikum, G. Fine-grained Semantic Typing of Emerging Entities **2013**. pp. 1488–1497.
234. Solovyev, V.; Gareev, R.; Ivanov, V. Dictionary and Pattern-Based Recognition of Organization Names in Russian News Texts. *Global Journal on Technology* **2013**, *3*.
235. Steinberger, R. Multilingual and Cross-Lingual News Analysis in the Europe Media Monitor (EMM) (Extended Abstract). In *Multidisciplinary Information Retrieval*; Hutchison, D.; Kanade, T.; Kittler, J.; Kleinberg, J.M.; Mattern, F.; Mitchell, J.C.; Naor, M.; Nierstrasz, O.; Pandu Rangan, C.; Steffen, B.; et al., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; Vol. 8201, pp. 1–4. [https://doi.org/10.1007/978-3-642-41057-4\\_1](https://doi.org/10.1007/978-3-642-41057-4_1).
236. Hoffart, J.; Altun, Y.; Weikum, G. Discovering emerging entities with ambiguous names. In Proceedings of the Proceedings of the 23rd international conference on World wide web - WWW '14, Seoul, Korea, 2014; pp. 385–396. <https://doi.org/10.1145/2566486.2568003>.
237. Kiritoshi, K.; Ma, Q. Named Entity Oriented Related News Ranking. In *Database and Expert Systems Applications*; Decker, H.; Lhotská, L.; Link, S.; Spies, M.; Wagner, R.R., Eds.; Springer International Publishing: Cham, 2014; Vol. 8645, pp. 82–96. [https://doi.org/10.1007/978-3-319-10085-2\\_7](https://doi.org/10.1007/978-3-319-10085-2_7).
238. Kompatsiaris, I.; Diplaris, S.; Papadopoulos, S. Social Data and Multimedia Analytics for News and Events Applications. In Proceedings of the EDBT/ICDT Workshops, 2014, pp. 280–281.
239. Vuurens, J.B.; de Vries, A.P.; Blanco, R.; Mika, P. Online News Tracking for Ad-Hoc Information Needs. In Proceedings of the Proceedings of the 2015 International Conference on Theory of Information Retrieval - ICTIR '15, Northampton, Massachusetts, USA, 2015; pp. 221–230. <https://doi.org/10.1145/2808194.2809474>.
240. Hou, L.; Li, J.; Wang, Z.; Tang, J.; Zhang, P.; Yang, R.; Zheng, Q. NewsMiner: Multifaceted news analysis for event search. *Knowledge-Based Systems* **2015**, *76*, 17–29. <https://doi.org/10.1016/j.knosys.2014.11.017>.
241. Shabat, H.A.; Omar, N. Named Entity Recognition in Crime News Documents Using Classifiers Combination. *Middle-East Journal of Scientific Research* **2015**, *23*, 1215–1221.
242. Verma, I.; Dey, L.; Srinivasan, R.S.; Singh, L. Event Detection from Business News. In *Pattern Recognition and Machine Intelligence*; Kryszkiewicz, M.; Bandyopadhyay, S.; Rybinski, H.; Pal, S.K., Eds.; Springer International Publishing: Cham, 2015; Vol. 9124, pp. 575–585. [https://doi.org/10.1007/978-3-319-19941-2\\_55](https://doi.org/10.1007/978-3-319-19941-2_55).
243. Li, H.; Fang, W.; An, H.; Huang, X. Words Analysis of Online Chinese News Headlines about Trending Events: A Complex Network Perspective. *PLOS ONE* **2015**, *10*, e0122174. <https://doi.org/10.1371/journal.pone.0122174>.
244. Segev, E. Visible and invisible countries: News flow theory revised. *Journalism: Theory, Practice & Criticism* **2015**, *16*, 412–428. <https://doi.org/10.1177/1464884914521579>.
245. Chakraborty, S.; Venkataraman, A.; Jagabathula, S.; Subramanian, L. Predicting Socio-Economic Indicators using News Events. In Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, San Francisco, California, USA, 2016; pp. 1455–1464. <https://doi.org/10.1145/2939672.2939817>.
246. Pino, C.; Kavasidis, I.; Spampinato, C. GeoSentiment: A tool for analyzing geographically distributed event-related sentiments. In Proceedings of the 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2016; pp. 270–271. <https://doi.org/10.1109/CCNC.2016.7444775>.
247. Rospocher, M.; van Erp, M.; Vossen, P.; Fokkens, A.; Aldabe, I.; Rigau, G.; Soroa, A.; Ploeger, T.; Bogaard, T. Building event-centric knowledge graphs from news. *Journal of Web Semantics* **2016**, *37-38*, 132–151. <https://doi.org/10.1016/j.websem.2015.12.004>.
248. Liu, B.; Niu, D.; Lai, K.; Kong, L.; Xu, Y. Growing story forest online from massive breaking news. In Proceedings of the Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017, pp. 777–785.
249. Lyu, L.; Fetahu, B. Real-time Event-based News Suggestion for Wikipedia Pages from News Streams. In Proceedings of the Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18, Lyon, France, 2018; pp. 1793–1799. <https://doi.org/10.1145/3184558.3191642>.
250. Petersen, S.E.; Ostendorf, M. Text Simplification for Language Learners: A Corpus Analysis. In Proceedings of the Workshop on Speech and Language Technology in Education, 2007.

251. Tseng, Y.H.; Chang, C.Y.; Rundgren, S.N.C.; Rundgren, C.J. Mining concept maps from news stories for measuring civic scientific literacy in media. *Computers & Education* **2010**, *55*, 165–177. <https://doi.org/10.1016/j.compedu.2010.01.002>.
252. Kovanović, V.; Joksimović, S.; Gašević, D.; Siemens, G.; Hatala, M. What public media reveals about MOOCs: A systematic analysis of news reports: What public media reveals about MOOCs. *British Journal of Educational Technology* **2015**, *46*, 510–527. <https://doi.org/10.1111/bjjet.12277>.
253. Montes-y Gómez, M.; Gelbukh, A.; López-López, A. A Statistical Approach to the Discovery of Ephemeral Associations among News Topics. In *Database and Expert Systems Applications*; Springer Berlin Heidelberg: Berlin, Heidelberg; Vol. 2113.
254. Berg, T.; Berg, A.; Edwards, J.; Maire, M.; White, R.; Yee-Whye Teh.; Learned-Miller, E.; Forsyth, D. Names and faces in the news. In Proceedings of the Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., Washington, DC, USA, 2004; Vol. 2, pp. 848–854. <https://doi.org/10.1109/CVPR.2004.1315253>.
255. Drury, B.; Almeida, J.J. Construction of a Local Domain Ontology from News Stories. In *Progress in Artificial Intelligence*; Lopes, L.S.; Lau, N.; Mariano, P.; Rocha, L.M., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; Vol. 5816, pp. 400–410. [https://doi.org/10.1007/978-3-642-04686-5\\_33](https://doi.org/10.1007/978-3-642-04686-5_33).
256. Feng, A.; Allan, J. Incident threading for news passages. In Proceedings of the Proceeding of the 18th ACM conference on Information and knowledge management - CIKM '09, Hong Kong, China, 2009; p. 1307. <https://doi.org/10.1145/1645953.1646118>.
257. Gwadera, R.; Crestani, F. Mining and ranking streams of news stories using cross-stream sequential patterns. In Proceedings of the Proceeding of the 18th ACM conference on Information and knowledge management - CIKM '09, Hong Kong, China, 2009; p. 1709. <https://doi.org/10.1145/1645953.1646210>.
258. Leetaru, K.; Schrodt, P.A. Gdelt: Global data on events, location, and tone, 1979–2012. In Proceedings of the ISA annual convention, 2013, Vol. 2, pp. 1–49.
259. Shi, B.; Ifrim, G.; Hurley, N. Insight4News: Connecting News to Relevant Social Conversations. In *Machine Learning and Knowledge Discovery in Databases*; Calders, T.; Esposito, F.; Hüllermeier, E.; Meo, R., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; Vol. 8726, pp. 473–476. [https://doi.org/10.1007/978-3-662-44845-8\\_38](https://doi.org/10.1007/978-3-662-44845-8_38).
260. Sardanios, C.; Katakis, I.M.; Petasis, G.; Karkaletsis, V. Argument Extraction from News. In Proceedings of the Proceedings of the 2nd Workshop on Argumentation Mining, Denver, CO, 2015; pp. 56–66. <https://doi.org/10.3115/v1/W15-0508>.
261. Cambi, F.; Crescenzi, P.; Pagli, L. Analyzing and comparing on-line news sources via (two-layer) incremental clustering. In Proceedings of the 8th International Conference on Fun with Algorithms, FUN 2016, 2016.
262. Ploch, D.; Lommatzsch, A.; Schultze, F. An Advanced Press Review System Combining Deep News Analysis and Machine Learning Algorithms. In Proceedings of the Proceedings of ACL-2016 System Demonstrations, Berlin, Germany, 2016; pp. 109–114. <https://doi.org/10.18653/v1/P16-4019>.
263. Rupnik, J.; Muhic, A.; Leban, G.; Skraba, P.; Fortuna, B.; Grobelnik, M. News Across Languages - Cross-Lingual Document Similarity and Event Tracking. *Journal of Artificial Intelligence Research* **2016**, *55*, 283–316. <https://doi.org/10.1613/jair.4780>.
264. Leppänen, L.; Munezero, M.; Sirén-Heikel, S.; Granroth-Wilding, M.; Toivonen, H. Finding and expressing news from structured data. In Proceedings of the Proceedings of the 21st International Academic Mindtrek Conference on - AcademicMindtrek '17, Tampere, Finland, 2017; pp. 174–183. <https://doi.org/10.1145/3131085.3131112>.
265. Liu, X.; Nourbakhsh, A.; Li, Q.; Shah, S.; Martin, R.; Duprey, J. Reuters Tracer: Toward Automated News Production Using Large Scale Social Media Data. *arXiv:1711.04068 [cs]* **2017**. arXiv: 1711.04068.
266. Hu, C.; Li, Y.; Wang, Y.; Wu, L. Analysis of Hot News Based on Big Data. In Proceedings of the 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS). IEEE, 2018, pp. 678–681.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.