EXPLORING SEASONAL TRENDS IN BULGARIAN MEDIA: INSIGHTS FROM A DATA-DRIVEN ANALYSIS OF WEB PUBLICATIONS

Plamen Milev[1]
e-mail: pmilev@unwe.bg [1]

Abstract

This paper presents a data-driven analysis of Bulgarian online media, aimed at identifying seasonal trends and event-driven fluctuations in news coverage. The study quantitatively analyzes the total monthly publication volume and the frequency of ten most popular keywords over a one-year period. The findings reveal two primary influences on the editorial agenda: predictable seasonal patterns, with a notable decrease in media output during the summer months, and the strong impact of specific high-profile events. For instance, coverage related to political terms like "Trump" and "Elections" peaked in direct alignment with major electoral cycles, while geopolitical topics such as "Ukraine" and "Israel" showed sharp fluctuations corresponding to key international developments. The paper discusses these dynamics, providing empirical insights into how the Bulgarian media agenda is shaped by the interplay of cyclical trends and event-driven news.

Key words: Web Scraping, Media Intelligence, Data Analysis, Web Publications, Editorial Focus

**JEL:** C88, L86.

Introduction

In the modern digital age, online media plays a crucial role in shaping public opinion and defining the public agenda. The constant 24-hour flow of information creates a dynamic and highly competitive environment where editorial priorities shift rapidly. Understanding the factors that influence these priorities is essential for analysts, researchers, and media professionals alike. While intuitive assumptions are often made about "calmer" or "busier" news periods, there is a lack of specific, data-driven research to outline these trends in the Bulgarian media context. This paper aims to fill this gap by conducting a quantitative analysis of web publications from leading Bulgarian media outlets. The primary objective of the research is to identify and analyze both recurring seasonal trends and specific event-driven fluctuations in media coverage. By measuring the total volume of publications and tracking the frequency of key topics, we aim to provide an empirical basis for understanding when and why certain topics dominate the news flow. For the purpose of this analysis, a full one-year period was examined – from April 2024 to March 2025. The methodology involves web scraping and processing of publication data from eight prominent media platforms: Bulgaria ON AIR, Bloomberg, BNR, BNT, bTV, DW, Eurocom, and Nova. The analysis includes two main components: first, an aggregated analysis of the total number of monthly publications to establish baseline levels of media activity and their seasonal fluctuations; and second, a detailed frequency analysis of the ten most frequent keywords identified through this scraping process. These prominent keywords ("Bulgaria," "Sofia," "Ukraine," "Trump," "USA," "Elections," "Russia," "Israel," "NATO," and "Europe") inherently represent a mix of domestic, geopolitical, and international political topics. As will be shown in the subsequent sections, the analysis reveals clear and predictable seasonal patterns, most notably a significant decline in media activity during the summer months. However, sharp, event-driven peaks stand out against this baseline. For example, coverage of topics related to the "USA" and "Trump" reaches its zenith in line with the US election cycle (November 2024 – January 2025), while topics like "Ukraine" and "Israel" show fluctuations directly related to the escalation of events. This paper proceeds by detailing the data collection methodology, followed by an analysis of the aggregated publication volumes to establish

[1] Associate Professor, PhD. Department of Information Technologies and Communications, University of National and World Economy, ORCID 0000-0002-4867-0586.

seasonal trends. Subsequently, it delves into the specific monthly dynamics of the top keywords, before concluding with a summary of the findings on the dual influence of seasonality and event-driven developments on the Bulgarian media agenda.

Methodology

The methodological approach of this paper is grounded in the contemporary challenges of digital data processing. The modern digital landscape is defined by the rapid integration of advanced technologies such as Artificial Intelligence [1] and novel user interfaces [2]. This technological shift necessitates new approaches for managing large-scale, cloud-based Big Data solutions [3] and, critically, for processing the vast amounts of unstructured data generated by web and social platforms [4]. This paper builds directly on this context. It specifically applies established principles of modern web data processing to handle the collection of online information [5]. Furthermore, it employs a systematic approach to organizing this information, aligning with core concepts of data cataloging [6]. To extract insights from the collected data, the research utilizes a quantitative approach that draws from several specific analytical methods. These include techniques for automated content analysis [7], intelligent tagging and search [8], natural language processing [9], and semantic analysis [10]. This combination of methods allows for the transformation of raw, unstructured web data into the meaningful, quantitative insights presented in this paper. Figure 1 illustrates the methodological approach of the research.
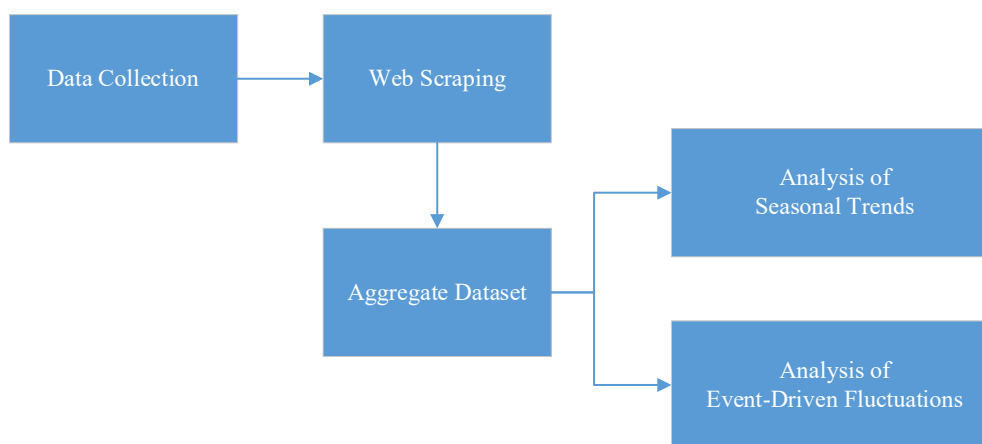
Source: Author's diagram illustrating the research process

Figure 1: The Methodological Flowchart

This paper employs a quantitative methodology based on web scraping and data analysis to examine publication trends in Bulgarian online media. The process is divided into two main stages: data collection and data analysis.

Data Collection

The dataset for this study was compiled by scraping publicly available information from eight prominent Bulgarian media platforms: Bulgaria ON AIR, Bloomberg, BNR, BNT, bTV, DW, Eurocom, and Nova. The data collection covers a complete one-year period, from April 1, 2024, to March 31, 2025, to ensure the inclusion of all seasonal variations.

The web scraping process was configured to extract the titles and publication timestamps of articles from the selected media sources. This focus on publication titles allows for a high-level analysis of the primary topics being presented to the public, as headlines are designed to encapsulate the most newsworthy aspect of a story.

Data Analysis

The collected raw data was processed and aggregated for analysis. The analysis phase consisted of two primary components:

- Aggregated publication analysis: the total number of publications from all eight media sources was aggregated on a monthly basis. This approach was used to establish a baseline for overall media output and to identify broad seasonal trends, such as peaks of activity or quieter periods (e.g., summer months, holidays).

- Keyword frequency analysis: the second component involved a detailed analysis of the publication titles. The entire corpus of collected titles was processed to identify the most frequently occurring and relevant terms. This data-driven approach resulted in the identification of the top 10 keywords: "Bulgaria," "Sofia," "Ukraine," "Trump," "USA," "Elections," "Russia," "Israel," "NATO," and "Europe." The monthly frequency of each of these keywords was then tracked individually to observe the dynamics of specific topics and their correlation with real-world events.

The application of this two-phase methodological approach – spanning from broad data collection to specific keyword frequency analysis – produced a granular dataset. The following section presents and discusses these results, detailing the identified seasonal patterns and event-driven fluctuations in the Bulgarian media's publication volume and thematic focus.

Results and Discussion

This section transitions from the outlined methodology to the empirical findings of the study. It presents a comprehensive analysis of the data collected from the eight selected Bulgarian media platforms. The analysis is structured to first provide a high-level overview of the aggregate monthly distributions to identify seasonal trends, followed by a granular analysis of the ten most frequent keywords, illustrating how editorial focus shifts in response to specific, event-driven national and international developments. This multi-layered approach provides a quantitative basis for understanding the dynamics of the Bulgarian digital news landscape.

Monthly Publication Trends

The initial data collection phase aggregated the total number of web publications from each of the eight selected media outlets for the 12-month period. This provided a baseline understanding of the relative output of each platform. To identify seasonal patterns, the total number of publications from all eight platforms was aggregated on a monthly basis (Table 1).

Table 1: Monthly Distribution of Publications

Month	Publications
April 2024	20526
May 2024	20210
June 2024	20444
July 2024	19936
August 2024	17736
September 2024	18144
October 2024	22037
November 2024	21125
December 2024	20224
January 2025	22740
February 2025	21632
March 2025	23368

Source: Author's data obtained from web scraping

The monthly distribution reveals a relatively consistent content generation throughout the year, with clear seasonal fluctuations. Noticeable peaks occur in March 2025 (23,368 publications) and January 2025 (22,740 publications), likely due to increased coverage related to significant political, economic, or social events common at the beginning of the year. Conversely, August 2024 registers the lowest activity (17,736 publications), reflecting typical seasonal patterns associated with vacation periods and reduced overall media output. September 2024 similarly shows relatively lower numbers, suggesting a gradual return to standard publication volumes following the summer season.

Keyword Fluctuation Analysis

The following section details the monthly frequency of the top 10 most popular keywords identified in the publication titles. This analysis highlights how editorial focus shifts in response to specific national and international events. Table 2 presents the monthly frequency of occurrences of the word "Bulgaria".

Table 2: Monthly Occurrences of "Bulgaria"

Month	Titles
April 2024	743
May 2024	756
June 2024	862
July 2024	649
August 2024	614
September 2024	863
October 2024	871
November 2024	709
December 2024	678
January 2025	691
February 2025	695
March 2025	877

Source: Author's data obtained from web scraping

The monthly distribution of "Bulgaria" reveals moderate fluctuation. Peaks appear in March 2025 (877 titles), October 2024 (871 titles), and September 2024 (863 titles), corresponding with significant domestic events. Lower frequencies are registered in the summer (July and August), reflecting a typical seasonal reduction in media coverage. Table 3 presents the monthly frequency of occurrences of the word "Sofia".

Table 3: Monthly Occurrences of "Sofia"

Month	Titles
April 2024	614
May 2024	802
June 2024	664
July 2024	438
August 2024	391
September 2024	621
October 2024	764
November 2024	687
December 2024	647
January 2025	691
February 2025	824
March 2025	633

Source: Author's data obtained from web scraping

The analysis of "Sofia" highlights clear variations. Significant peaks are observable in February 2025 (824 titles) and May 2024 (802 titles), suggesting increased coverage of events in the capital. Noticeably lower coverage occurred during the summer (July and August), following a similar seasonal pattern to "Bulgaria". Table 4 presents the monthly frequency of occurrences of the word "Ukraine".

Table 4: Monthly Occurrences of "Ukraine"

Month	Titles
April 2024	541
May 2024	577
June 2024	589
July 2024	478
August 2024	351
September 2024	370
October 2024	397
November 2024	647
December 2024	499
January 2025	360
February 2025	1006
March 2025	1003

Source: Author's data obtained from web scraping

The frequency of "Ukraine" displays considerable variability. A significant peak is evident in February 2025 (1,006 titles) and March 2025 (1,003 titles), which could be linked to critical geopolitical or military developments. Conversely, a noticeable drop appears during the late summer, suggesting decreased media emphasis. Table 5 presents the monthly frequency of occurrences of the word "Trump".

Table 5: Monthly Occurrences of "Trump"

Month	Titles
April 2024	90
May 2024	113
June 2024	92
July 2024	552
August 2024	186
September 2024	268
October 2024	243
November 2024	1034
December 2024	341
January 2025	1151
February 2025	1191
March 2025	1113

Source: Author's data obtained from web scraping

The frequency of "Trump" strongly correlates with political events. A remarkable increase occurs from November 2024 (1,034 titles) through February 2025 (1,191 titles). This period coincides with critical political events, most likely connected to the US presidential elections and their aftermath. Table 6 presents the monthly frequency of occurrences of the word "USA".

Table 6: Monthly Occurrences of "USA"

Month	Titles
April 2024	452
May 2024	435
June 2024	358
July 2024	373
August 2024	346
September 2024	307
October 2024	412
November 2024	645
December 2024	409
January 2025	717
February 2025	855
March 2025	895

Source: Author's data obtained from web scraping

The distribution of "USA" exhibits noticeable variability, with distinct peaks toward the end of 2024 and the beginning of 2025. Increased coverage is observed from November 2024 (645 titles) to March 2025 (895 titles), reflecting heightened media attention driven by major political events, including the US presidential elections and post-election dynamics. Table 7 presents the monthly frequency of occurrences of the word "Elections".

Table 7: Monthly Occurrences of "Elections"

Month	Titles
April 2024	566
May 2024	390
June 2024	622
July 2024	303
August 2024	254
September 2024	522
October 2024	669
November 2024	629
December 2024	191
January 2025	165
February 2025	197
March 2025	263

Source: Author's data obtained from web scraping

The distribution of "Elections" shows patterns related to political events. Notable peaks are evident in October 2024 (669 titles) and November 2024 (629 titles), likely coinciding with the US presidential elections. Another increase is observed in June 2024 (622 titles), a period which may reflect regional or European political events. Markedly lower coverage from December 2024 onwards suggests a post-election period. Table 8 presents the monthly frequency of occurrences of the word "Russia".

Table 8: Monthly Occurrences of "Russia"

Month	Titles
April 2024	537
May 2024	293
June 2024	198
July 2024	133
August 2024	203

September 2024	342
October 2024	707
November 2024	271
December 2024	124
January 2025	213
February 2025	128
March 2025	190

Source: Author's data obtained from web scraping

The occurrences of "Russia" exhibit moderate fluctuations. Notable peaks are evident in March 2025 (580 titles) and November 2024 (448 titles), likely correlating with significant geopolitical developments. Lower frequencies appear during the summer (July and August), reflecting reduced editorial emphasis. Table 9 presents the monthly frequency of occurrences of the word "Israel".

Table 9: Monthly Occurrences of "Israel"

Month	Titles
April 2024	537
May 2024	293
June 2024	198
July 2024	133
August 2024	203
September 2024	342
October 2024	707
November 2024	271
December 2024	124
January 2025	213
February 2025	128
March 2025	190

Source: Author's data obtained from web scraping

The monthly occurrences of "Israel" show substantial fluctuations. The most prominent peaks appear in October 2024 (707 titles) and April 2024 (537 titles), likely indicating major political, military, or diplomatic events. Coverage declines sharply during other months, reaching its lowest points in December 2024 and February 2025. Table 10 presents the monthly frequency of occurrences of the word "NATO".

Table 10: Monthly Occurrences of "NATO"

Month	Titles
April 2024	208
May 2024	263
June 2024	300
July 2024	261
August 2024	255
September 2024	184
October 2024	247
November 2024	280
December 2024	266
January 2025	329
February 2025	280
March 2025	280

Source: Author's data obtained from web scraping

The frequency of "NATO" remains relatively stable, with moderate fluctuations. Slight peaks are evident in January 2025 (329 titles) and June 2024 (300 titles), possibly connected to NATO summits or significant policy announcements. The lowest frequency occurs in September 2024 (184 titles). Table 11 presents the monthly frequency of occurrences of the word "Europe".

Table 11: Monthly Occurrences of "Europe"

Month	Titles
April 2024	180
May 2024	333
June 2024	230
July 2024	185
August 2024	144
September 2024	212
October 2024	236
November 2024	279
December 2024	187
January 2025	266
February 2025	407
March 2025	424

Source: Author's data obtained from web scraping

The occurrences of "Europe" demonstrate moderate fluctuations. A significant increase occurs in February 2025 (407 titles) and March 2025 (424 titles), indicating increased media interest in European Union activities or regional developments. Another peak in May 2024 (333 titles) may correspond to EU-related events. The lowest number is seen in August 2024.

In summary, the presented data analysis confirms a clear model of dual influence on the Bulgarian online media agenda. On one hand, the aggregate publication data demonstrates a distinct and predictable seasonal cycle, characterized by a significant decline in media activity during the late summer period (August). On the other hand, the detailed keyword frequency analysis reveals a high degree of reactivity to specific, high-impact events. This is most evident in the sharp peaks in coverage for topics like "Trump" and "USA," correlating directly with the U.S. election cycle, and the surges in mentions of "Ukraine" and "Israel," coinciding with key geopolitical developments. These findings collectively illustrate the dynamic interplay between predictable, seasonal rhythms and the unpredictable, event-driven nature of the modern news cycle.

Conclusion

This paper provides empirical evidence that the Bulgarian digital news agenda operates under a dual influence: it is anchored by a predictable seasonal cycle but is simultaneously and powerfully steered by the unpredictable cadence of major national and global events. Future research could expand upon this methodology by moving beyond title-based keyword analysis to full-text sentiment analysis, or by comparing these established media trends against the corresponding discourse on social media platforms to identify potential disparities or time lags in the public conversation.

Acknowledgement

References

1. Dimitrov, K. (2025). The Growing Role of VR And AI Industry to Transform Soft Skills Development. *Trakia Journal of Sciences*, *23*(2), 298-303.

2. Mihova, V. (2023). Expanding The Possibilities of Information System for Working with Voice. *Innovative Information Technologies for Economy Digitalization (IITED)*, (1), 152-157.
3. Kovacheva, M. (2025). Evaluating Challenges and Opportunities Associated with Cloud-Based Big Data Solutions. *Nauchni trudove*, (2), 175-188.
4. Velkova, I. (2023). Unstructured social media data processing with artificial intelligence. *Industry 4.0*, *8*(2), 65-67.
5. Tabov, Y. (2024). Web Data Processing in the Digital Age: Challenges and Solutions. *Innovative Information Technologies for Economy Digitalization (IITED)*, (1), 223-228.
6. Stefanov, G. (2023). Data Cataloging in Big Data Environment. *Innovative Information Technologies for Economy Digitalization (IITED)*, (1), 102-107.
7. Belev, I. (2024). Combining digital business automation technologies. *Innovative Information Technologies for Economy Digitalization (IITED)*, (1), 305-312.
8. Marzovanova, M. (2015). Intelligent Tagging and Search as a Fully Automated System. *Economic Alternatives*, (2), 95-100.
9. Radoev, M. (2024). Application of Natural Language Processing in Accounting. *Innovative Information Technologies for Economy Digitalization (IITED)*, (1), 107-113.
10. Lazarova, V. (2025). Artificial Intelligence as a Service. *Ikonomiceski i Sotsialni Alternativi*, (2), 128-141.