

Article

Not peer-reviewed version

Mapping the Digital Media Landscape in Bulgaria: Analysis of Web Publications

[Plamen Hristov Milev](#) *

Posted Date: 7 April 2025

doi: 10.20944/preprints202504.0574.v1

Keywords: media intelligence; web publications; natural language processing; data analytics; web scraping



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.

Article

Mapping the Digital Media Landscape in Bulgaria: Analysis of Web Publications

Plamen Milev

Department of Information Technologies and Communications, University of National and World Economy,
1700 Sofia, Bulgaria; pmilev@unwe.bg; Tel.: +35928195312

Abstract: This study explores the thematic structure and editorial focus of the digital media landscape in Bulgaria by analyzing one year of online news publications from eight major media outlets. The data were collected through a custom-built web scraping application developed in Java, which enabled the automated extraction and processing of full-text articles from publicly accessible news websites. The structured dataset, generated during the scraping process, records word-level occurrences in both article titles and bodies, along with publication dates and URLs. By applying lexical frequency analysis and temporal tracking, the study identifies the most frequently used words, their monthly distribution, and platform-specific usage patterns. The findings reveal clear distinctions in editorial focus between public broadcasters, private national media, and international outlets. Additionally, the analysis highlights how title construction and word prominence vary depending on platform type and media strategy. The study demonstrates the potential of web scraping and computational text analysis as scalable tools for investigating media systems in small and transitional democracies.

Keywords: media intelligence; web publications; natural language processing; data analytics; web scraping

1. Introduction

The exponential growth of digital content and the transformation of traditional media into dynamic, multi-platform ecosystems have fundamentally reshaped the way societies produce, distribute, and consume information. In this rapidly evolving context, understanding the thematic structure and editorial dynamics of digital media has become crucial for researchers, policymakers, and media professionals alike. Understanding how media platforms construct thematic focus and editorial priorities has long been a central concern in communication research. Foundational theories such as agenda-setting and framing have highlighted the influence of media in shaping public perception and political discourse [1]. These concepts remain relevant in the digital era, where editorial choices are increasingly mediated by algorithms, audience metrics, and content dynamics. The comparative model of media systems developed by Hallin and Mancini provides a useful lens for comparing editorial strategies, particularly in countries like Bulgaria that exhibit features of the polarized pluralist model [2]. In the Bulgarian context, digital media landscape has undergone complex transformations in recent years, particularly regarding pluralism and access to information [3]. As news increasingly migrates to online environments, questions emerge about how media platforms prioritize topics, frame narratives, and balance domestic with international coverage. Over the past decade, scholars have proposed various approaches to studying digital news production – ranging from qualitative discourse analysis to large-scale computational techniques. While qualitative methods offer depth, they often struggle with scalability and temporal coverage. In contrast, data-driven approaches such as natural language processing (NLP) and web scraping enable the automated analysis of large volumes of unstructured media data, offering new opportunities for real-time and longitudinal media research [4–6]. Scientific repositories, enriched

through web scraping, can substantially improve literature searches and bibliometric analyses [7,8]. However, despite the increasing availability of computational tools, much of the existing research has focused on dominant international outlets or English-language media, often overlooking smaller national media systems or emerging democracies. Moreover, questions remain about how different types of media – public, private, international – diverge in their editorial agendas, and how they respond to political, social, and geopolitical developments. Some studies suggest convergence toward a common global news agenda, while others highlight persistent fragmentation and national editorial priorities [9,10]. This study seeks to contribute to this ongoing debate by analyzing the content of eight major Bulgarian media outlets – encompassing public broadcasters, private national platforms, and internationally affiliated source – using a combination of web scraping and NLP techniques. The aim is to uncover thematic trends, identify differences in editorial focus, and better understand how digital media in a smaller European country navigates both local and global information flows. The research highlights clear distinctions in how different platforms address topics such as national governance, elections, international conflicts, and geopolitical alliances. By focusing on word frequency, title emphasis, and monthly dynamics, the study offers a granular look into the structure of digital media content over the course of one year (April 2024 – March 2025). These findings underscore the importance of media intelligence approaches in studying news ecosystems and offer empirical insights relevant not only to media studies, but also to political science, communication, and digital sociology.

2. Materials and Methods

To collect and process the dataset used in this study, a custom web scraping Java desktop application was developed. The system combines JavaFX for user interface management and embedded browser functionality with jsoup for robust HTML parsing and content extraction. Web scraping has proven highly effective in various fields, including analyzing consumer sentiment in tourism [10], gathering structured data for film rating analytics [11], decentralizing and verifying news sources using blockchain technology [12], and extracting key phrases from web content [13]. The application targets eight major Bulgarian media websites: Bulgaria ON AIR, Bloomberg TV Bulgaria, BNR, BNT, bTV, DW (Bulgarian service), Eurocom, and Nova TV – selected for their broad audience reach and continuous digital publishing activity. The application employs the JavaFX components WebView and WebEngine to simulate a lightweight browser capable of loading both static and dynamically generated web content [14]. The internal logic and interaction between the system's components are illustrated in Figure 1.

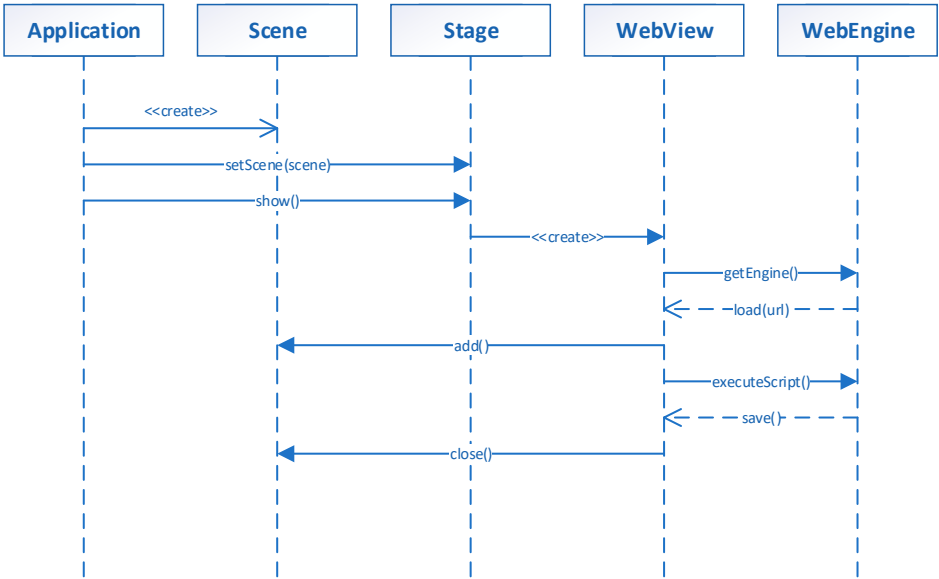


Figure 1. Sequence diagram showing the interaction between JavaFX components during the initialization and web content processing workflow.

Upon launching, the Application creates a Stage and displays a Scene containing a WebView, which loads the specified media URL. The associated WebEngine executes the page rendering and, where needed, runs JavaScript scripts to ensure that dynamically generated elements are fully accessible. Once the webpage is fully loaded, its HTML content is extracted. This raw HTML is then passed to the jsoup parser, which systematically analyzes and extracts key article components, such as the title, publication date, main body text, and page metadata [15]. A key characteristic of the system is that it performs full-text lexical analysis of every article during runtime. Rather than relying on a predefined set of keywords, the application extracts all words found in both the article title and body. For each word encountered, the system records whether it appears in the title, in the text, or in both. In addition, the structured output includes the publication date and the URL of each article. No full-text articles were stored after processing. Once the analytical extraction was complete, only the structured dataset – containing word occurrences and associated metadata – was retained. This design ensures efficient data handling and minimizes long-term storage requirements, while enabling the generation of statistically sound lexical datasets. This data model also ensures compliance with web usage guidelines. The resulting structured dataset served as the foundation for determining the most frequently occurring words over the one-year period (April 2024 – March 2025) and supported further analyses on temporal trends and editorial focus across media platforms. This approach aligns with contemporary practices in computational content analysis, where full-text data are used to extract structured insights at scale [16].

The selected media platforms reflect a diversity of ownership structures, editorial profiles, and audience orientations. Bulgaria ON AIR is a national media outlet combining television broadcasting with an active online presence, covering both domestic affairs and international developments [17]. Bloomberg TV Bulgaria is a business-focused platform affiliated with the global Bloomberg network, providing financial news, economic commentary, and global market insight [18]. BNR (Bulgarian National Radio) is Bulgaria’s public radio broadcaster, offering comprehensive coverage of political, economic, and cultural developments, with an emphasis on public service content [19]. BNT (Bulgarian National Television) serves as the country’s national public television channel, delivering wide-ranging domestic news and institutional reporting [20]. bTV is a leading private television network in Bulgaria, offering mainstream news coverage with a blend of local and international focus

[21]. DW (Deutsche Welle, Bulgarian edition) is the Bulgarian-language service of Germany’s international broadcaster, with strong emphasis on foreign policy, European affairs, and global analysis [22]. Eurocom is a private Bulgarian television channel with an independent editorial stance, known for its political commentary and attention to domestic issues [23]. Nova TV is a major commercial broadcaster in Bulgaria, offering dynamic coverage of current events, with a mix of national reporting and coverage of international developments [24]. These sources were selected to ensure a representative overview of Bulgaria’s digital media environment, encompassing both public and private, national and international perspectives.

3. Results

The selection of media sources included in this analysis was based on their significant influence, audience reach, and active digital presence within the Bulgarian media landscape. Specifically, the selected platforms are prominent media outlets with television and/or radio coverage, representing a mixture of public and private media, as well as local and international news providers. This approach ensures a diverse and comprehensive overview of online media content in Bulgaria. The analyzed dataset covers a period of one year, specifically from April 2024 to March 2025. This timeframe was chosen to provide an up-to-date and representative snapshot of contemporary media coverage, capturing seasonal trends, significant national and international events, and editorial dynamics across the examined platforms. The data collected from these sources are presented in the following tables and graphically illustrated in corresponding figures to facilitate clear and insightful analysis.

Table 1 presents the total number of publications for each of the analyzed media platforms within the period from April 2024 to March 2025.

Table 1. Total Number of Publications by Media Platform (April 2024–March 2025).

Media	Website	Publications
Bulgaria ON AIR	https://bgonair.bg	34213
Bloomberg	https://bloombergtv.bg	13842
BNR	https://bnr.bg/	52692
BNT	https://bntnews.bg	41520
bTV	https://btvnovinite.bg	25911
DW	https://www.dw.com/bg/	2719
Eurocom	https://eurocom.bg	32358
Nova	https://nova.bg/news	44867

Figure 2 provides a graphical representation of the data shown in Table 1.

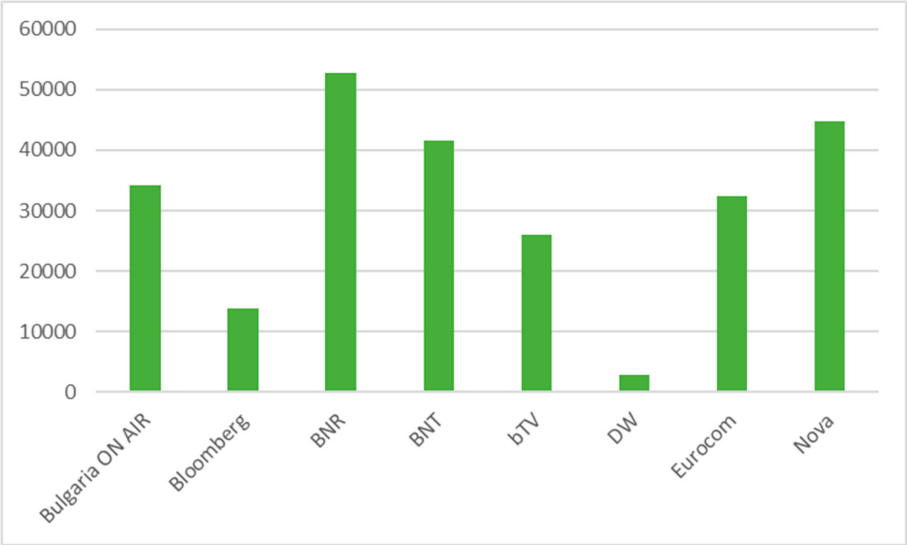


Figure 2. Graphical Representation of the Total Number of Publications by Media Platform (April 2024–March 2025).

The data clearly illustrates considerable variations in the quantity of published content across the examined media platforms. The Bulgarian National Radio (BNR) recorded the highest total number of publications, amounting to 52,692, followed by Nova (44,867 publications) and the Bulgarian National Television (BNT) with 41,520 publications. These three media outlets evidently play leading roles in terms of news coverage and information dissemination within the country. Platforms such as Bulgaria ON AIR (34,213 publications) and Eurocom (32,358 publications) also demonstrate considerable activity, highlighting their significance within the Bulgarian digital media landscape. Other platforms like bTV (25,911 publications) and Bloomberg (13,842 publications) show moderate content activity. In stark contrast, Deutsche Welle (DW) published significantly fewer articles – only 2,719. This disparity is likely due to DW's specific editorial strategy focusing on selected international and socially significant topics, as opposed to local Bulgarian platforms, which tend to provide more frequent daily news coverage. These initial findings offer an essential starting point for subsequent analyses, helping to identify the dominant information sources in Bulgaria's digital media ecosystem and enabling the formulation of hypotheses regarding editorial policies and target audiences of the various media platforms.

Table 2 presents the number of publications aggregated monthly across all analyzed media platforms, covering the period from April 2024 to March 2025.

Table 2. Monthly Distribution of Publications (April 2024–March 2025).

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

Figure 3 provides a graphical representation of the data shown in Table 2.

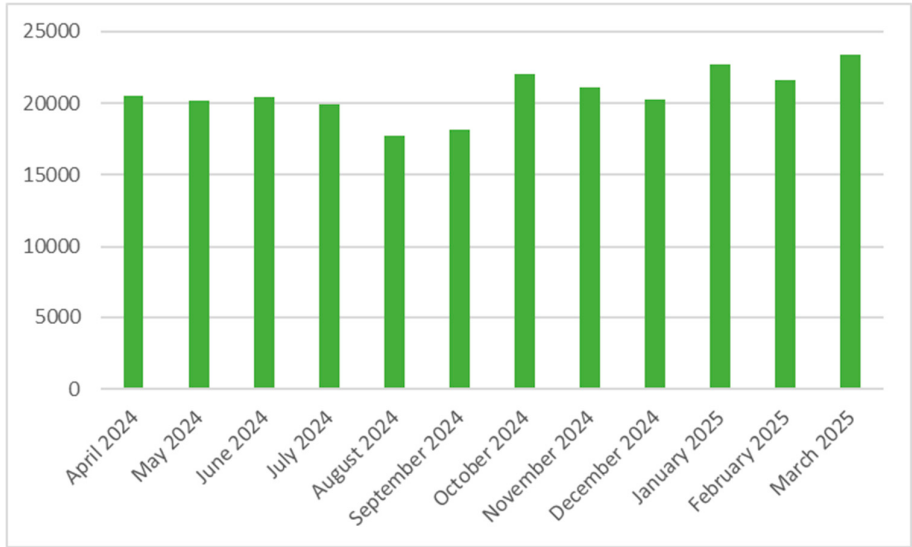


Figure 3. Graphical Representation of the Monthly Distribution of Publications (April 2024–March 2025).

The monthly distribution of publications 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. October 2024 also shows heightened media activity with 22,037 publications, potentially reflecting increased editorial attention around national and international developments during this period. 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. Overall, these monthly variations underline the responsiveness of media platforms to seasonal influences, holidays, and significant events, reflecting adaptive editorial strategies that align with audience interests and newsworthiness.

Table 3 presents the 20 most frequently occurring words in publication titles, along with their occurrences in the main body of the publication texts, during the analyzed period (April 2024–March 2025).

Table 3. Top 20 Most Frequent Words in Publication Titles and Corresponding Text Occurrences (April 2024–March 2025).

Word	Occurrences in Titles	Occurrences in Text
Bulgaria	9008	55816
Sofia	7776	30777
Ukraine	6818	21113
Trump	6374	15106
USA	6204	29274
Elections	4771	24628
Russia	4603	18176
Israel	3339	9480
NATO	3153	6248
Europe	3083	20880
Government	2629	26030
Germany	2198	11616

Greece	2100	7174
Gaza	1998	6373
China	1983	9304
Turkey	1801	7179
Paris	1662	7104
France	1480	9472
Iran	1294	4744
Romania	1283	5855

Figure 4 provides a graphical representation of the data shown in Table 3 in the context of the top 20 most frequent words in publication titles.

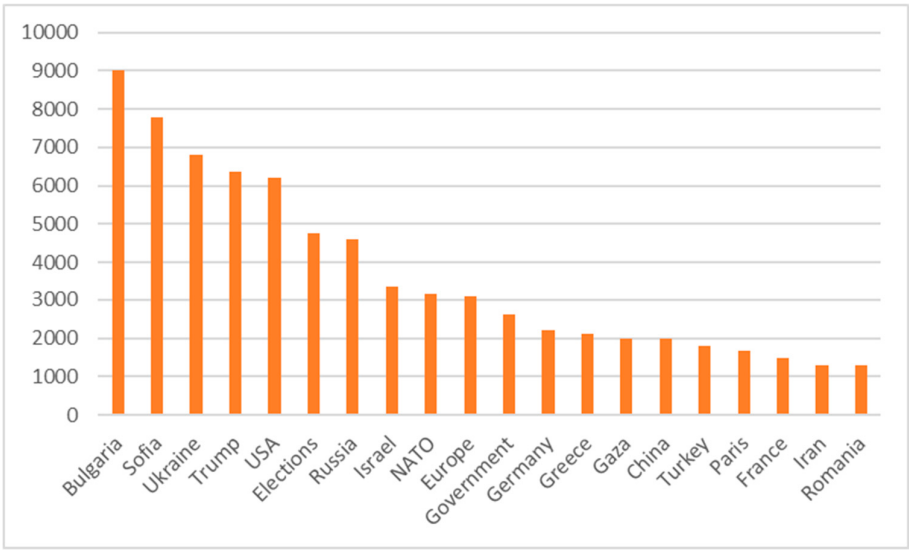


Figure 4. Graphical Representation of the Top 20 Most Frequent Words in Publication Titles (April 2024–March 2025).

The analysis of the most frequently used words in publication titles highlights key thematic focuses of the studied media platforms. The prominence of geographic references such as "Bulgaria" (9,008 occurrences) and "Sofia" (7,776 occurrences) underlines a strong emphasis on domestic news and local events. Frequent mentions of countries and geopolitical terms including "Ukraine," "USA," "Russia," "Israel," "Germany," "Greece," "China," "Turkey," "France," "Iran," and "Romania" reflect significant editorial attention to international affairs, geopolitical tensions, and diplomatic relations. The presence of "Trump" (6,374 occurrences) and thematic words such as "Elections" and "Government" indicate an ongoing focus on politics and governance issues. The occurrence of "Paris" (1,662 occurrences) is likely driven by heightened media attention around the upcoming 2024 Summer Olympic Games hosted in the French capital.

Figure 5 provides a graphical representation of the data shown in Table 3 in the context of the occurrences of the top 20 title words in publication texts.

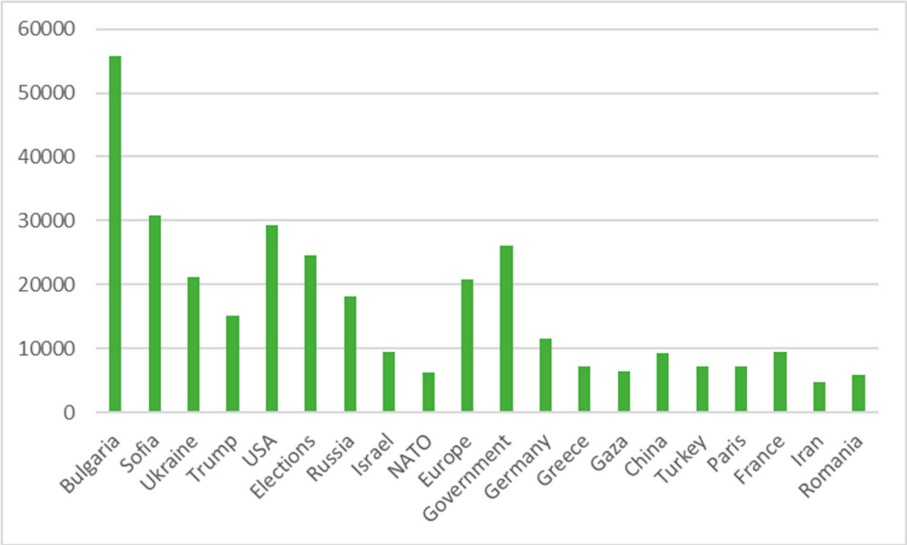


Figure 5. Graphical Representation of the Occurrences of the Top 20 Title Words in Publication Texts (April 2024–March 2025).

The frequency of these title words in publication texts further validates their thematic significance within media content. Words such as "Bulgaria" (55,816 occurrences) and "Sofia" (30,777 occurrences) maintain their prominence, underscoring continuous coverage of domestic topics throughout the articles. Terms like "Government" (26,030), "Elections" (24,628), and "USA" (29,274) highlight in-depth discussions related to political and governance issues, both domestically and internationally. The extensive mentions of geopolitical entities such as "Ukraine," "Russia," and "Europe" demonstrate consistent editorial interest in global affairs. Lower but significant frequencies of terms such as "Paris" (7,104 occurrences) further reflect sustained interest related to internationally relevant events, notably the Paris 2024 Olympics, generating deeper textual coverage beyond headlines. It is noteworthy that the ranking of words based on their frequency in publication titles does not completely align with their frequency within the publication texts. Such discrepancies suggest differences in editorial strategies between selecting attention-grabbing headlines and delivering detailed content within the articles. For instance, certain terms like "Trump," "NATO," and "Israel" have relatively high prominence in titles compared to their frequency in article texts, likely due to their newsworthiness and capacity to attract readers’ attention. Conversely, words such as "Government," "Europe," and "Bulgaria," which have higher occurrences within texts, indicate deeper thematic coverage beyond mere headlines. This observation underscores the distinct roles played by titles (aiming to capture reader interest quickly) and article bodies (intended to provide comprehensive and detailed information).

To better understand the consistency and variations in editorial attention toward prominent topics over the studied period, we conducted a monthly analysis of the top 10 most frequent words identified previously. This approach allows us to highlight any substantial fluctuations or seasonal patterns in media focus on specific issues, entities, or geographical references.

Table 4 presents the monthly frequency of occurrences of the word "Bulgaria" in publication titles from April 2024 to March 2025.

Table 4. Monthly Occurrences of the Word "Bulgaria" in Publication Titles (April 2024–March 2025).

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

Figure 6 provides a graphical representation of the data shown in Table 4.

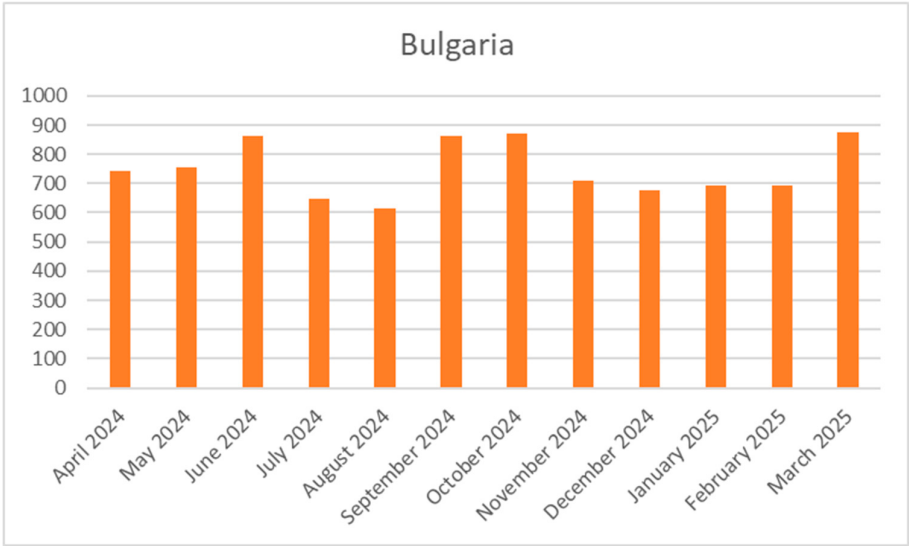


Figure 6. Graphical Representation of the Monthly Occurrences of the Word "Bulgaria" in Publication Titles (April 2024–March 2025).

The monthly distribution of the occurrences of "Bulgaria" in publication titles reveals a moderate but notable fluctuation throughout the analyzed period. Peaks in the usage of "Bulgaria" appear during the months of March 2025 (877 titles), October 2024 (871 titles), September 2024 (863 titles), and June 2024 (862 titles), potentially corresponding with significant domestic political, economic, or social events prompting increased national coverage. Conversely, lower frequencies of the word were registered in the summer months – July (649 titles) and August (614 titles) – reflecting a typical seasonal reduction in media coverage, possibly due to decreased political and economic activity during the vacation period. Overall, despite observable fluctuations, the word "Bulgaria" consistently remains a central and frequently used keyword, underscoring its stable and persistent importance within the editorial content of analyzed media platforms.

Table 5 presents the monthly frequency of occurrences of the word "Sofia" in publication titles from April 2024 to March 2025.

Table 5. Monthly Occurrences of the Word "Sofia" in Publication Titles (April 2024–March 2025).

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

Figure 7 provides a graphical representation of the data shown in Table 5.

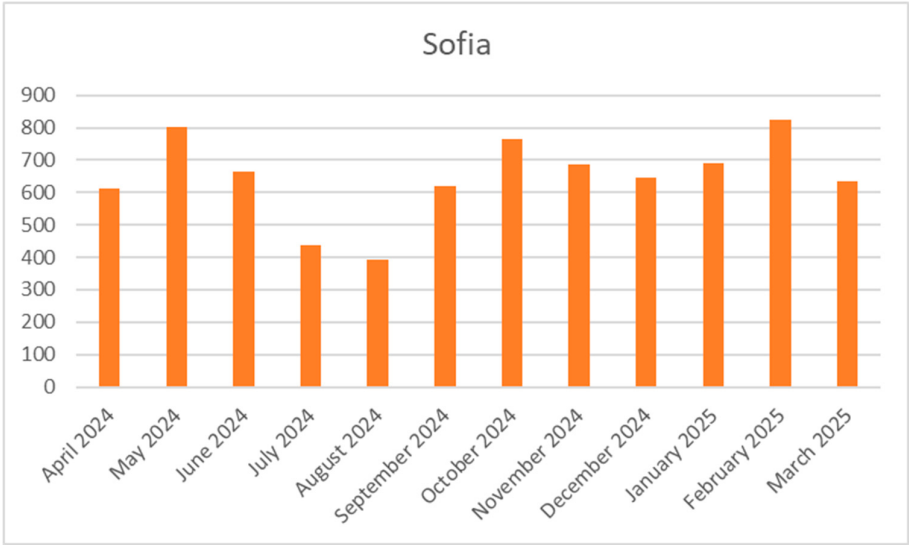


Figure 7. Graphical Representation of the Monthly Occurrences of the Word "Sofia" in Publication Titles (April 2024–March 2025).

The monthly analysis of the term "Sofia" in publication titles highlights clear variations in editorial attention throughout the studied period. Significant peaks are observable in February 2025 (824 titles) and May 2024 (802 titles), suggesting increased media coverage possibly linked to important political, economic, or social events centered in the capital city during these periods. Noticeably lower coverage occurred during the summer months, particularly in July 2024 (438 titles) and August 2024 (391 titles), following a similar seasonal pattern to that observed for "Bulgaria," reflecting decreased urban and institutional activity during these months. The distribution indicates that while the focus on Sofia as a central geographical and political reference is consistently high, specific events likely dictate fluctuations in the city's prominence in headlines over the year.

Table 6 presents the monthly frequency of occurrences of the word "Ukraine" in publication titles from April 2024 to March 2025.

Table 6. Monthly Occurrences of the Word "Ukraine" in Publication Titles (April 2024–March 2025).

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

Figure 8 provides a graphical representation of the data shown in Table 6.

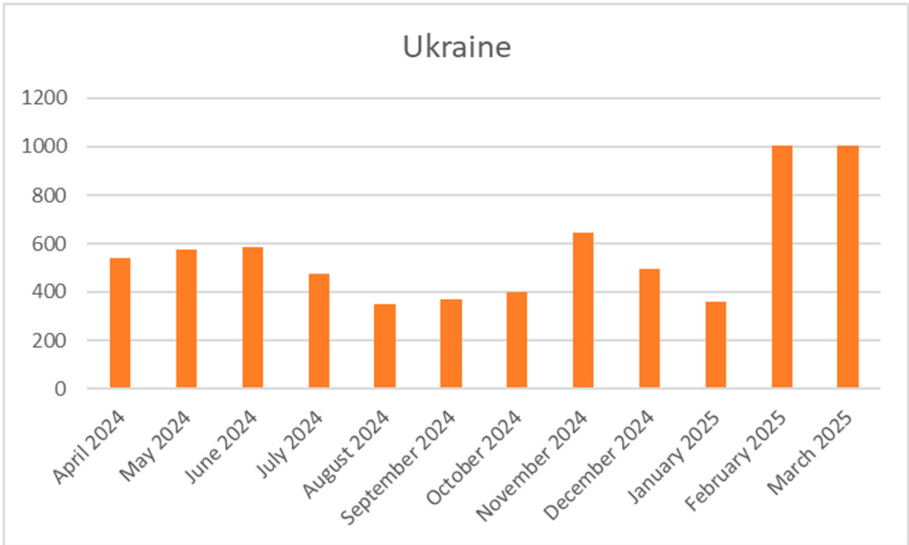


Figure 8. Graphical Representation of the Monthly Occurrences of the Word "Ukraine" in Publication Titles (April 2024–March 2025).

The monthly frequency of the word "Ukraine" in publication titles displays considerable variability, reflecting shifting editorial priorities tied closely to geopolitical developments. A significant peak in coverage is evident in February 2025 (1,006 titles) and March 2025 (1003 titles). These substantial increases could be linked to critical geopolitical or military developments involving Ukraine, possibly including intensified conflicts, diplomatic negotiations, or other major international events attracting heightened media attention. Conversely, a noticeable drop in occurrences appears during the late summer months, particularly in August 2024 (351 titles) and September 2024 (370 titles), suggesting decreased media emphasis, potentially due to a temporary reduction in newsworthy events related to Ukraine during this period. Overall, these data indicate that the coverage related to Ukraine is significantly event-driven, aligning closely with developments in the geopolitical landscape and influencing the editorial agenda across the studied Bulgarian media outlets.

Table 7 presents the monthly frequency of occurrences of the word "Trump" in publication titles from April 2024 to March 2025.

Table 7. Monthly Occurrences of the Word "Trump" in Publication Titles (April 2024–March 2025).

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

Figure 9 provides a graphical representation of the data shown in Table 7.

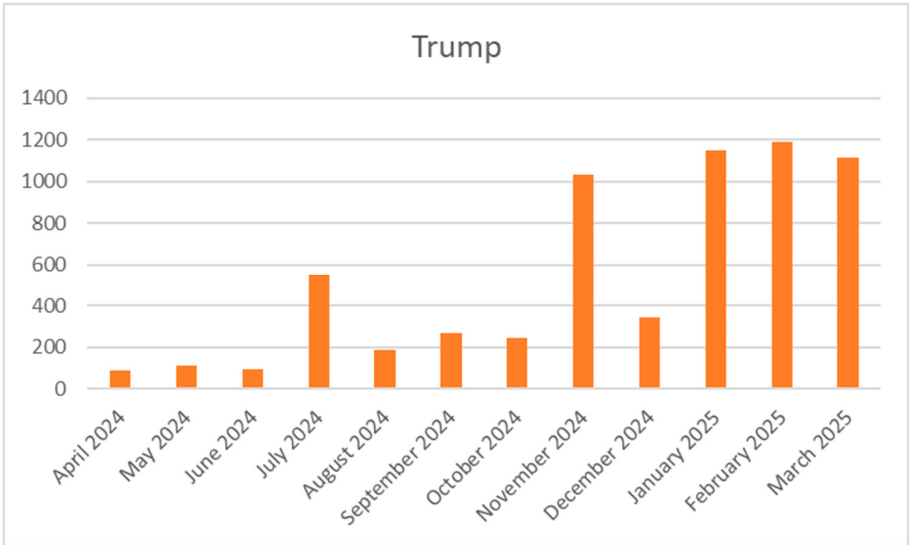


Figure 9. Graphical Representation of the Monthly Occurrences of the Word "Trump" in Publication Titles (April 2024–March 2025).

The frequency of the word "Trump" in publication titles reveals significant fluctuations, strongly correlating with notable political events and developments associated with former US President Donald Trump. A remarkable increase occurs from November 2024 (1,034 titles) through February 2025 (1,191 titles), with a peak in media coverage. This period coincides with critical political events, most likely connected to the US presidential elections, electoral campaigns, or significant political controversies involving Donald Trump. In contrast, markedly lower occurrences were recorded from April to June 2024, suggesting relatively lower media interest or fewer newsworthy events concerning Trump during this timeframe. The sharp spikes from late 2024 into early 2025 emphasize the extent to which media coverage of international figures like Trump is heavily event-driven and influenced by political cycles, elections, and significant controversies or developments.

Table 8 presents the monthly frequency of occurrences of the word "USA" in publication titles from April 2024 to March 2025.

Table 8. Monthly Occurrences of the Word "USA" in Publication Titles (April 2024–March 2025).

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

Figure 10 provides a graphical representation of the data shown in Table 8.

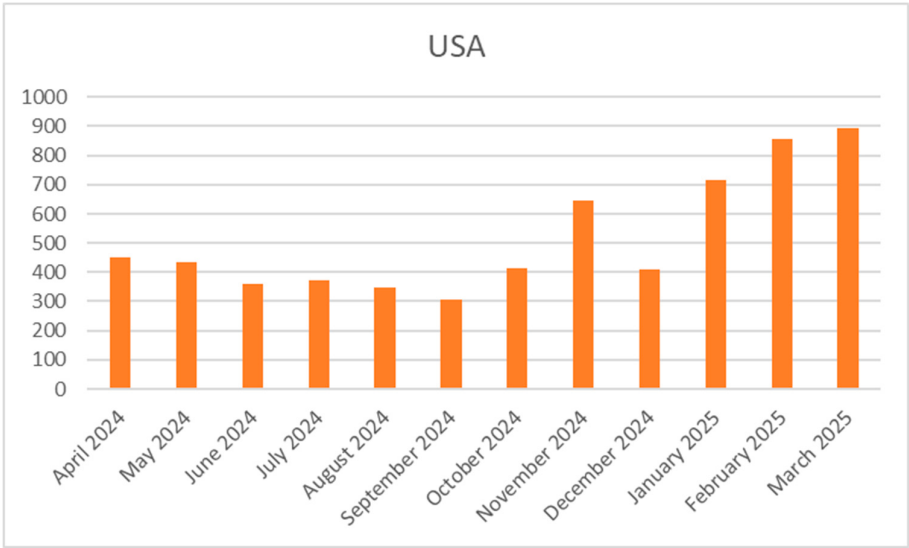


Figure 10. Graphical Representation of the Monthly Occurrences of the Word "USA" in Publication Titles (April 2024–March 2025).

The monthly distribution of the word "USA" in publication titles exhibits noticeable variability, with distinct peaks toward the end of 2024 and the beginning of 2025. Significantly increased coverage is observed from November 2024 (645 titles) to March 2025 (895 titles), reflecting heightened media attention possibly driven by major political events, including the U.S. presidential elections, shifts in foreign policy, or prominent international developments involving the United States. The highest peak in March 2025 (895 titles) and similarly high frequencies in February (855 titles) and January 2025 (717 titles) further align with post-election coverage, transition periods, and related political dynamics. Lower occurrences from June to September 2024 suggest a period of relatively subdued media interest, likely reflecting fewer significant newsworthy events involving the USA during these months. Overall, the observed monthly variations highlight how Bulgarian media coverage related to the USA closely mirrors the dynamics of political cycles and international events involving American political and economic influence.

Table 9 presents the monthly frequency of occurrences of the word "Elections" in publication titles from April 2024 to March 2025.

Table 9. Monthly Occurrences of the Word "Elections" in Publication Titles (April 2024–March 2025).

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

Figure 11 provides a graphical representation of the data shown in Table 9.

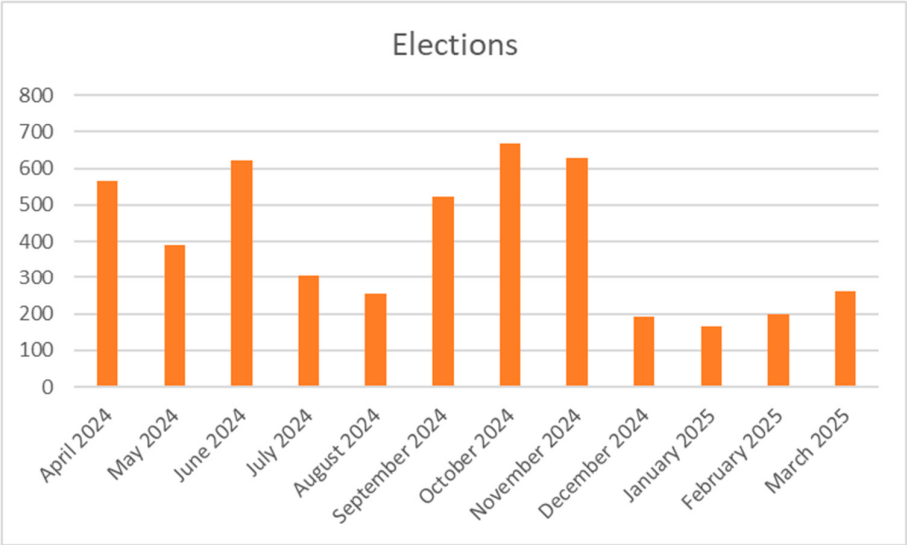


Figure 11. Graphical Representation of the Monthly Occurrences of the Word "Elections" in Publication Titles (April 2024–March 2025).

The monthly distribution of the word "Elections" in publication titles shows distinct patterns related to political events occurring during the analyzed period. Notable peaks in media coverage are evident in October 2024 (669 titles) and November 2024 (629 titles), likely coinciding with significant national or international election events, including coverage of the U.S. presidential elections and potential local or European political elections. Another prominent increase is observed earlier in June 2024 (622 titles), possibly reflecting regional or European political events or campaigns drawing considerable media attention. Markedly lower media coverage from December 2024 to March 2025 (ranging between 165 and 263 titles) suggests a period of decreased electoral activities and fewer election-related developments, typical for post-election periods and early calendar months. Overall, the fluctuations clearly illustrate how media attention toward elections is inherently tied to specific electoral cycles and political timelines, demonstrating peaks around major voting events and significant decreases during less politically active months.

Table 10 presents the monthly frequency of occurrences of the word "Russia" in publication titles from April 2024 to March 2025.

Table 10. Monthly Occurrences of the Word "Russia" in Publication Titles (April 2024–March 2025).

Month	Titles
April 2024	332
May 2024	464
June 2024	341
July 2024	296
August 2024	293
September 2024	337
October 2024	391
November 2024	448
December 2024	390
January 2025	289
February 2025	442
March 2025	580

Figure 12 provides a graphical representation of the data shown in Table 10.

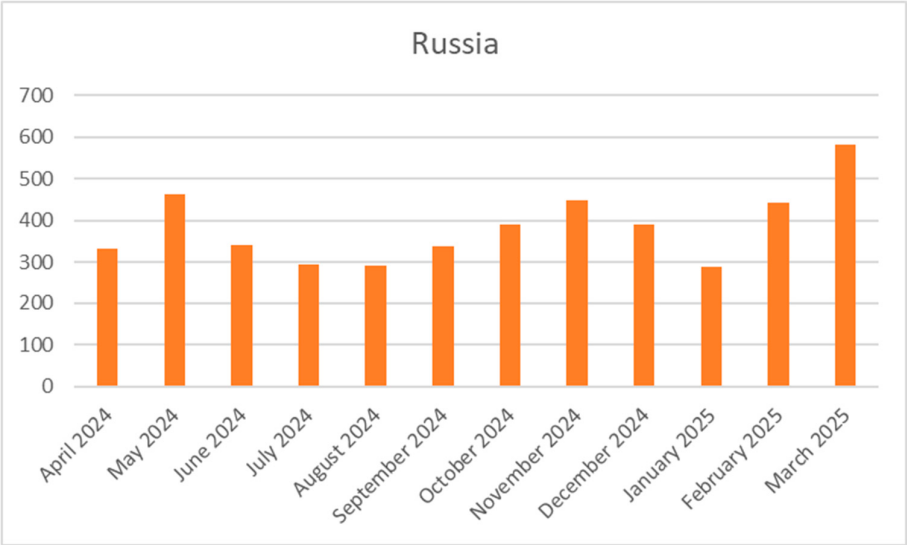


Figure 12. Graphical Representation of the Monthly Occurrences of the Word "Russia" in Publication Titles (April 2024–March 2025).

The monthly occurrences of the word "Russia" in publication titles exhibit moderate fluctuations throughout the analyzed period. Notable peaks in coverage are evident in March 2025 (580 titles) and November 2024 (448 titles), likely correlating with significant geopolitical developments, diplomatic interactions, or international events involving Russia, potentially connected to regional conflicts, energy policy issues, or diplomatic relations. Conversely, lower frequencies appear during the summer months, particularly July 2024 (296 titles) and August 2024 (293 titles), reflecting reduced editorial emphasis possibly due to fewer notable events or reduced political activities involving Russia during these periods. Overall, media coverage related to Russia remains consistent but is evidently responsive to significant international and regional events, highlighting its ongoing relevance within Bulgarian media coverage and global geopolitics.

Table 11 presents the monthly frequency of occurrences of the word "Israel" in publication titles from April 2024 to March 2025.

Table 11. Monthly Occurrences of the Word "Israel" in Publication Titles (April 2024–March 2025).

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

Figure 13 provides a graphical representation of the data shown in Table 11.

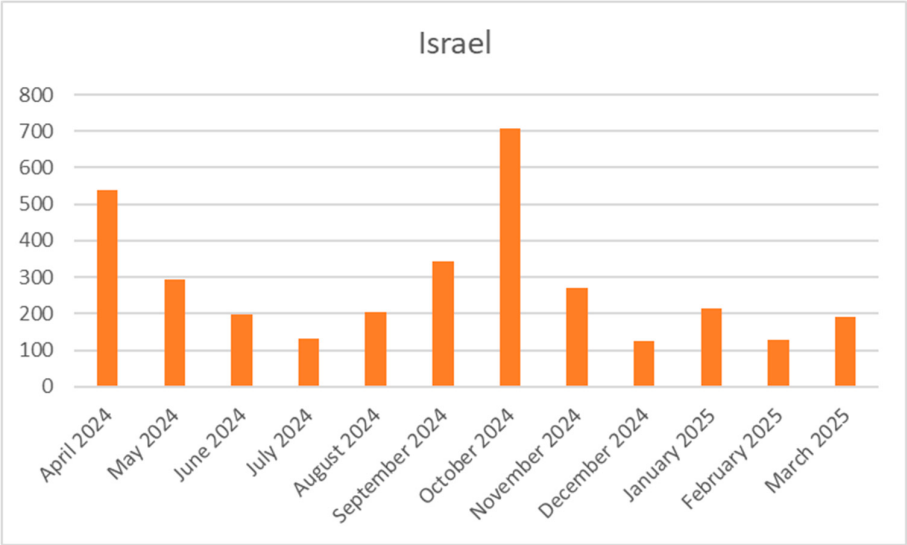


Figure 13. Graphical Representation of the Monthly Occurrences of the Word "Israel" in Publication Titles (April 2024–March 2025).

The monthly occurrences of the word "Israel" show substantial fluctuations, reflecting the event-driven nature of media coverage surrounding the country. The most prominent peaks appear in October 2024 (707 titles) and April 2024 (537 titles), likely indicating major political, military, or diplomatic events involving Israel, such as significant geopolitical tensions, conflicts, or high-profile diplomatic activities. Coverage declines sharply during other months, reaching its lowest points in December 2024 (124 titles) and July 2024 (133 titles), suggesting a period of relatively reduced newsworthy activity involving Israel during these months. Overall, the coverage of Israel appears strongly influenced by specific events, crises, or diplomatic interactions, resulting in notable peaks interspersed with periods of lower media visibility.

Table 12 presents the monthly frequency of occurrences of the word "NATO" in publication titles from April 2024 to March 2025.

Table 12. Monthly Occurrences of the Word "NATO" in Publication Titles (April 2024–March 2025).

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

Figure 14 provides a graphical representation of the data shown in Table 12.

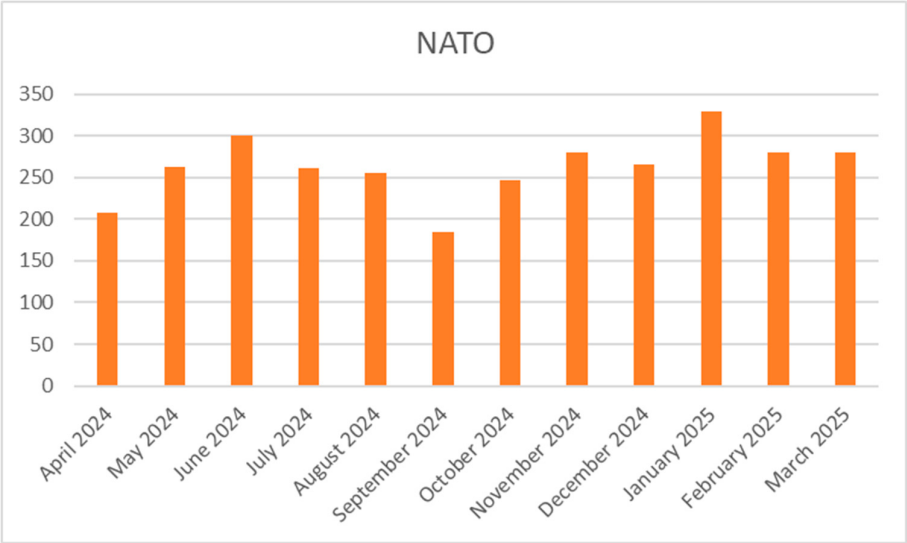


Figure 14. Graphical Representation of the Monthly Occurrences of the Word "NATO" in Publication Titles (April 2024–March 2025).

The frequency of the word "NATO" in publication titles remains relatively stable across the studied period, with moderate fluctuations indicating a steady media focus on NATO-related topics. Slight peaks are evident in January 2025 (329 titles) and June 2024 (300 titles), possibly connected to NATO summits, significant policy announcements, or strategic developments involving the alliance. The lowest frequency occurs in September 2024 (184 titles), reflecting a temporary decline in relevant NATO news or activity during this month. Nonetheless, coverage remains generally consistent, reflecting NATO's ongoing relevance in international security, geopolitical discussions, and Bulgaria's own strategic positioning within the alliance. Overall, these data illustrate that NATO-related coverage is driven primarily by scheduled or emergent geopolitical and security events rather than exhibiting extreme fluctuations.

Table 13 presents the monthly frequency of occurrences of the word "Europe" in publication titles from April 2024 to March 2025.

Table 13. Monthly Occurrences of the Word "Europe" in Publication Titles (April 2024–March 2025).

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

Figure 15 provides a graphical representation of the data shown in Table 13.

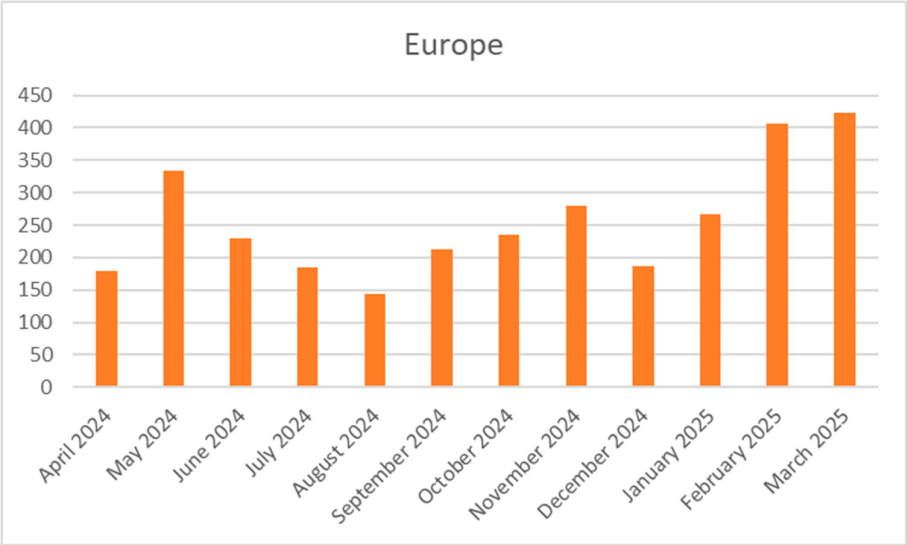


Figure 15. Graphical Representation of the Monthly Occurrences of the Word "Europe" in Publication Titles (April 2024–March 2025).

The monthly occurrences of the word "Europe" in publication titles demonstrate moderate fluctuations throughout the analyzed period. A significant increase occurs in February 2025 (407 titles) and remains relatively high in March 2025 (424 titles), indicating potentially increased media interest in European Union activities, economic policies, diplomatic interactions, or regional developments during this timeframe. Another noticeable peak in May 2024 (333 titles) may correspond to important EU-related events or discussions, such as European parliamentary activities, economic updates, or geopolitical decisions affecting the continent. The lowest number of occurrences is seen in August 2024 (144 titles), reflecting a typically lower media activity during summer months when fewer significant political or institutional activities occur in Europe. Overall, media coverage of "Europe" appears consistently present but clearly responsive to key European political, economic, and diplomatic events or developments, thus reflecting periodic increases in editorial attention toward the continent.

To achieve a deeper understanding of editorial priorities and thematic preferences among individual media platforms, we conducted a detailed analysis of how frequently the top 20 identified keywords appear in each platform's publications. This analysis reveals distinct editorial strategies, varying focuses on domestic and international topics, and potential audience targeting differences across media outlets.

Table 14 presents the distribution of top 20 words in publications of Bulgaria ON AIR from April 2024 to March 2025.

Table 14. Distribution of Top 20 Words in Publications of Bulgaria ON AIR (April 2024–March 2025).

Word	Publications
Bulgaria	7866
Sofia	4026
Ukraine	2666
Trump	1362
USA	2896
Elections	2790
Russia	2290
Israel	1184
NATO	635
Europe	2089

Government	2760
Germany	959
Greece	1260
Gaza	733
China	828
Turkey	844
Paris	670
France	851
Iran	509
Romania	571

Figure 16 provides a graphical representation of the data shown in Table 14.

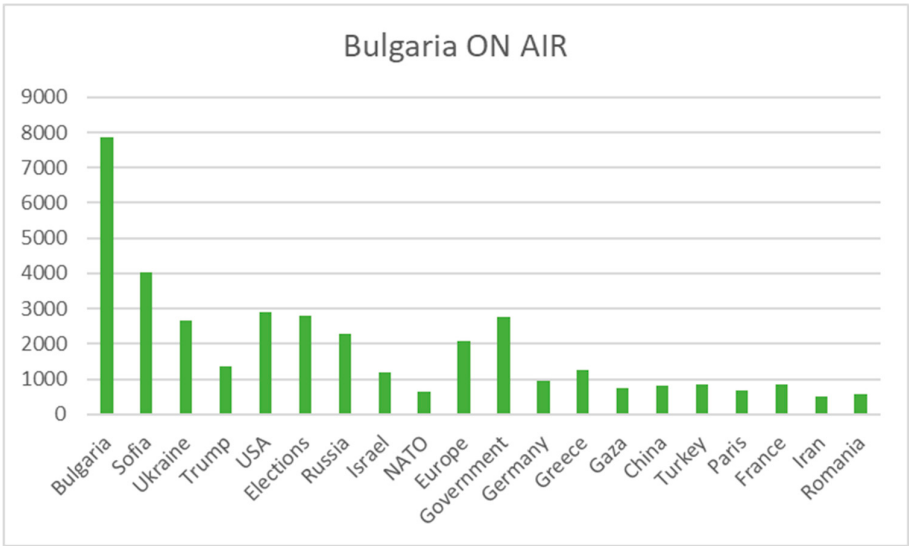


Figure 16. Graphical Representation of the Distribution of Top 20 Words in Publications of Bulgaria ON AIR (April 2024–March 2025).

The keyword distribution in the content of Bulgaria ON AIR indicates a strong editorial interest in both domestic and international themes, with a clear balance between local relevance and global reporting. The dominant presence of “Bulgaria” (7,866) and “Sofia” (4,026) confirms consistent attention to national and urban issues, likely reflecting daily political, social, and infrastructural developments. Additionally, “Government” (2,760) and “Elections” (2,790) highlight the channel’s sustained focus on political processes and institutional coverage. Among international topics, terms like “USA” (2,896), “Ukraine” (2,666), “Russia” (2,290), and “Europe” (2,089) appear frequently, reflecting global political tensions and regional diplomacy – likely driven by ongoing conflicts and EU affairs. This suggests a relatively even distribution of coverage between Western and Eastern geopolitical interests. Notably, words associated with Middle Eastern conflicts – such as “Israel”, “Gaza”, “Iran”, and “Turkey” – also maintain a significant presence, pointing to a comprehensive international news agenda. Coverage of “Paris” (670) and “France” (851) may be related to high-profile global events like the 2024 Olympics or EU developments. In summary, Bulgaria ON AIR demonstrates a content strategy that intertwines domestic political reporting with timely international developments, positioning itself as a news source with both local grounding and global awareness.

Table 15 presents the distribution of top 20 words in publications of Bloomberg from April 2024 to March 2025.

Table 15. Distribution of Top 20 Words in Publications of Bloomberg (April 2024–March 2025).

Word	Publications
Bulgaria	1341
Sofia	247
Ukraine	2505
Trump	3742
USA	7320
Elections	2368
Russia	2485
Israel	1248
NATO	740
Europe	3813
Government	3313
Germany	1635
Greece	269
Gaza	3891
China	921
Turkey	526
Paris	399
France	1207
Iran	980
Romania	265

Figure 17 provides a graphical representation of the data shown in Table 15.

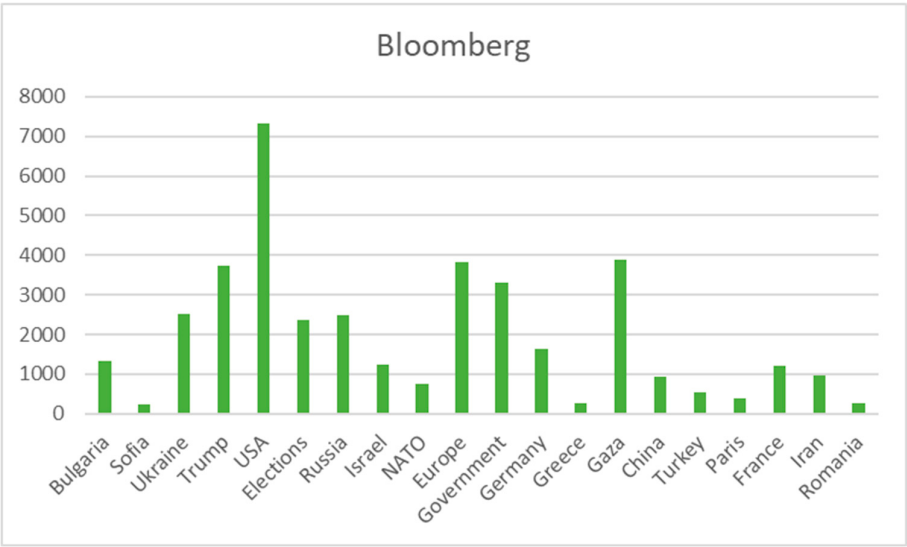


Figure 17. Graphical Representation of the Distribution of Top 20 Words in Publications of Bloomberg (April 2024–March 2025).

The data for Bloomberg TV Bulgaria reaffirm the outlet’s international and policy-oriented focus, with minimal emphasis on local content. Terms like “USA” (7,320), “Gaza” (3,891), “Trump” (3,742), “Europe” (3,813), and “Government” (3,313) are among the most prominent, reflecting deep editorial engagement with global political developments, international conflict, and leadership dynamics. Although “Bulgaria” (1,341) and “Sofia” (247) are present, their relatively low frequencies suggest limited coverage of national or local topics, aligning with Bloomberg’s global positioning and target audience. This contrast is especially noticeable when compared to public broadcasters such as

BNR or BNT. The strong presence of “Russia”, “Ukraine”, “China”, “Germany”, and “France” reflects focused attention on major global actors and regions relevant to economic and geopolitical analysis. The especially high number of mentions of “Gaza” (even surpassing “Trump”) indicates extensive coverage of Middle Eastern conflicts – likely viewed through the lenses of global stability, humanitarian impact, and international policy response. Topics related to elections (2,368) and governance (3,313) further emphasize the outlet’s interest in institutional processes, democratic systems, and global leadership transitions – core themes for Bloomberg’s economic-political reporting model. In summary, Bloomberg TV Bulgaria stands out for its strongly international, policy-centric editorial strategy, with limited domestic coverage but deep engagement with global developments that affect markets, security, and international governance.

Table 16 presents the distribution of top 20 words in publications of BNR from April 2024 to March 2025.

Table 16. Distribution of Top 20 Words in Publications of BNR (April 2024–March 2025).

Word	Publications
Bulgaria	13176
Sofia	8149
Ukraine	4625
Trump	2467
USA	5071
Elections	5400
Russia	3798
Israel	1941
NATO	1356
Europe	4632
Government	5389
Germany	2751
Greece	2024
Gaza	1573
China	1368
Turkey	1560
Paris	1622
France	2315
Iran	821
Romania	1564

Figure 18 provides a graphical representation of the data shown in Table 16.

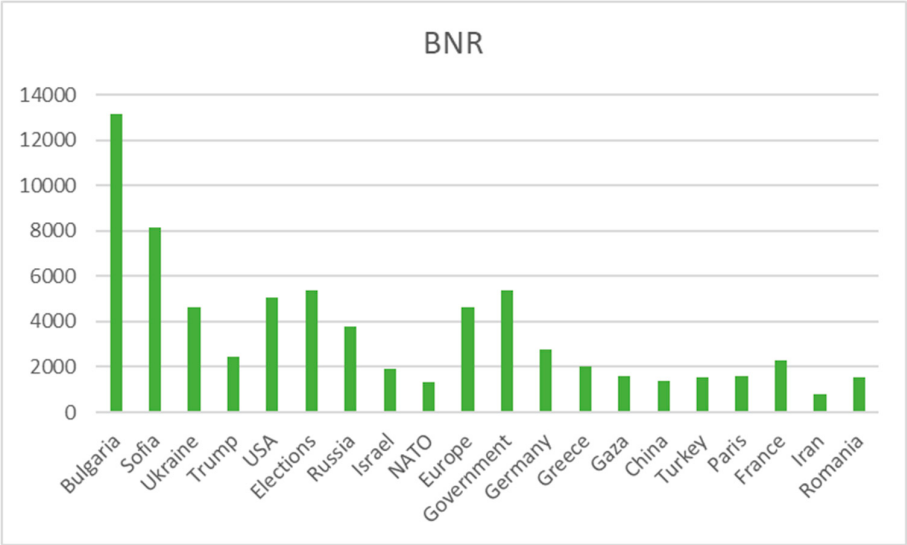


Figure 18. Graphical Representation of the Distribution of Top 20 Words in Publications of BNR (April 2024–March 2025).

The data for BNR confirms its profile as a public service broadcaster with a strong focus on national and institutional topics. The most frequently occurring words – “Bulgaria” (13,176), “Sofia” (8,149), “Elections” (5,400), and “Government” (5,389) – reflect the outlet’s consistent commitment to informing the public about domestic affairs, governance, and political processes. The frequency of “Ukraine” (4,625), “Russia” (3,798), “USA” (5,071), and “Europe” (4,632) indicates substantial attention to global developments, especially those with geopolitical significance for Bulgaria and the European Union. This supports BNR’s role not only as a national broadcaster, but also as a gateway to international information for the Bulgarian audience. Mentions of “NATO” (1,356), “Israel” (1,941), and “Gaza” (1,573) reflect coverage of international security, military alliances, and conflicts – themes that are both timely and relevant in the current global context. At the same time, regional countries such as Greece, Romania, and Turkey appear with relatively high frequency, confirming a regional orientation in addition to global and national reporting. The coverage of “Trump” (2,467), “France” (2,315), and “Germany” (2,751) points to consistent attention to the political dynamics in key Western countries, likely in relation to diplomacy, European affairs, or global leadership shifts. In conclusion, BNR maintains a well-balanced editorial strategy, with a strong core of national reporting complemented by robust international coverage. The data highlight BNR’s role as a comprehensive and reliable source of both domestic and global news for Bulgarian audiences.

Table 17 presents the distribution of top 20 words in publications of BNT from April 2024 to March 2025.

Table 17. Distribution of Top 20 Words in Publications of BNT (April 2024–March 2025).

Word	Publications
Bulgaria	10294
Sofia	5628
Ukraine	2325
Trump	1188
USA	3109
Elections	3174
Russia	1927
Israel	1347
NATO	732
Europe	2884

Government	2992
Germany	2111
Greece	1294
Gaza	832
China	725
Turkey	1326
Paris	2500
France	2128
Iran	590
Romania	1245

Figure 19 provides a graphical representation of the data shown in Table 17.

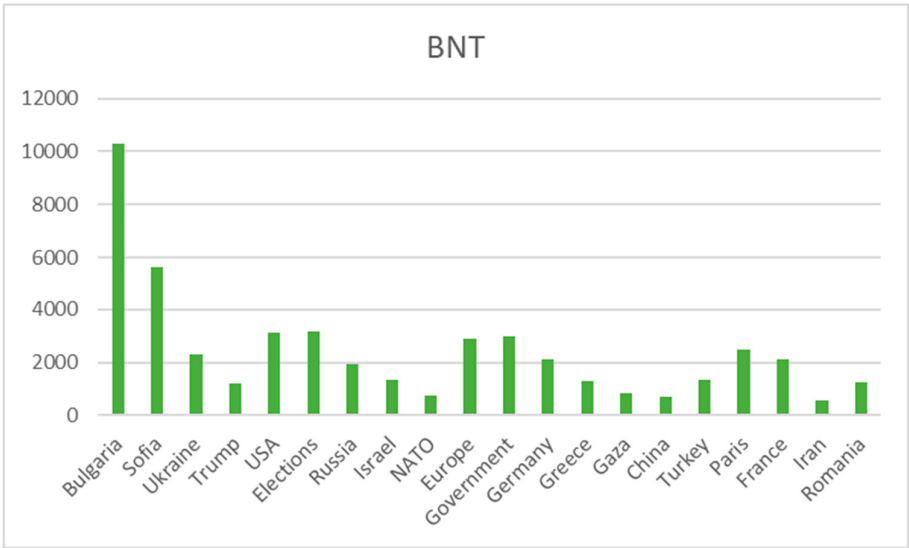


Figure 19. Graphical Representation of the Distribution of Top 20 Words in Publications of BNT (April 2024–March 2025).

The lexical distribution for BNT, Bulgaria’s national public television broadcaster, demonstrates a balanced editorial approach, combining extensive national coverage with regular attention to international developments. The high frequency of domestic terms such as “Bulgaria” (10,294) and “Sofia” (5,628) reflects the outlet’s public service mission to inform citizens about national and municipal issues. Similarly, the substantial use of “Elections” (3,174) and “Government” (2,992) confirms BNT’s role in reporting on institutional processes, democratic governance, and policy matters. While the international terms appear in lower absolute numbers compared to some other platforms, they still indicate solid global engagement. The presence of “USA” (3,109), “Europe” (2,884), “Ukraine” (2,325), and “Russia” (1,927) shows that BNT maintains consistent coverage of global geopolitical developments, particularly those affecting the region and Bulgaria’s international position. Mentions of “Germany” (2,111), “France” (2,128), and especially “Paris” (2,500) suggest coverage of European Union affairs and major global events – likely including the 2024 Summer Olympics in Paris, which could explain the notable prominence of the city name. Terms such as “Israel”, “Gaza”, and “Iran” reflect coverage of Middle Eastern conflicts, albeit at a more moderate level than in international or business-focused outlets. The regional perspective is reinforced through frequent references to Greece, Turkey, and Romania. Overall, BNT’s content profile aligns with its role as a national broadcaster – informing the public on local governance and national politics, while maintaining a clear, though proportionally smaller, window to the international stage.

Table 18 presents the distribution of top 20 words in publications of bTV from April 2024 to March 2025.

Table 18. Distribution of Top 20 Words in Publications of bTV (April 2024–March 2025).

Word	Publications
Bulgaria	6542
Sofia	4069
Ukraine	2093
Trump	1465
USA	2761
Elections	2839
Russia	1836
Israel	993
NATO	642
Europe	1947
Government	2869
Germany	1013
Greece	764
Gaza	606
China	608
Turkey	878
Paris	423
France	768
Iran	509
Romania	654

Figure 20 provides a graphical representation of the data shown in Table 18.

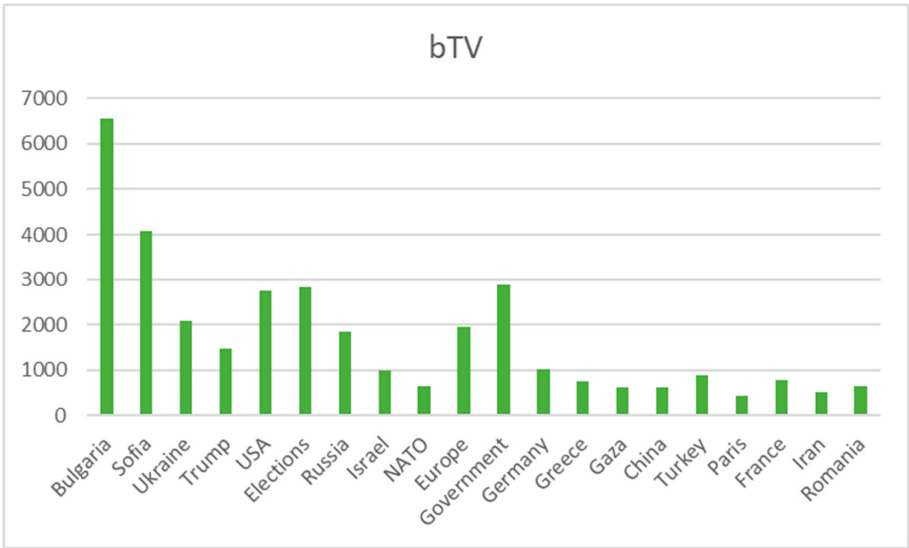


Figure 20. Graphical Representation of the Distribution of Top 20 Words in Publications of bTV (April 2024–March 2025).

The content distribution for bTV, one of Bulgaria’s largest private national broadcasters, reveals a balanced editorial profile with a slightly stronger domestic orientation complemented by solid international reporting. High frequencies of “Bulgaria” (6,542) and “Sofia” (4,069) confirm consistent coverage of national and local events. The prominence of “Government” (2,869) and “Elections”

(2,839) indicates a strong focus on political processes, public institutions, and electoral dynamics – key themes in national reporting. Although the frequencies for international terms are not as high as in internationally oriented platforms like Bloomberg or DW, bTV maintains a well-rounded foreign affairs section. Terms such as “USA” (2,761), “Ukraine” (2,093), “Russia” (1,836), and “Europe” (1,947) demonstrate attention to major global developments, particularly those affecting Bulgaria’s geopolitical context. Coverage of Middle Eastern issues is present but relatively modest, with “Israel”, “Gaza”, “Iran”, and “NATO” each appearing in under 1,000 publications. This suggests that international conflict is covered selectively, likely based on event salience or relevance to domestic discourse. Regional focus is visible through mentions of Greece (764), Turkey (878), and Romania (654), suggesting attention to Bulgaria’s neighboring countries and EU partners. In conclusion, bTV exhibits a dual editorial focus, combining consistent reporting on national politics and institutions with contextual international coverage, without showing the extreme thematic skew typical of either public broadcasters or international news agencies.

Table 19 presents the distribution of top 20 words in publications of DW from April 2024 to March 2025.

Table 19. Distribution of Top 20 Words in Publications of DW (April 2024–March 2025).

Word	Publications
Bulgaria	600
Sofia	187
Ukraine	827
Trump	511
USA	946
Elections	674
Russia	835
Israel	263
NATO	301
Europe	722
Government	978
Germany	815
Greece	114
Gaza	271
China	196
Turkey	223
Paris	89
France	246
Iran	173
Romania	128

Figure 21 provides a graphical representation of the data shown in Table 19.

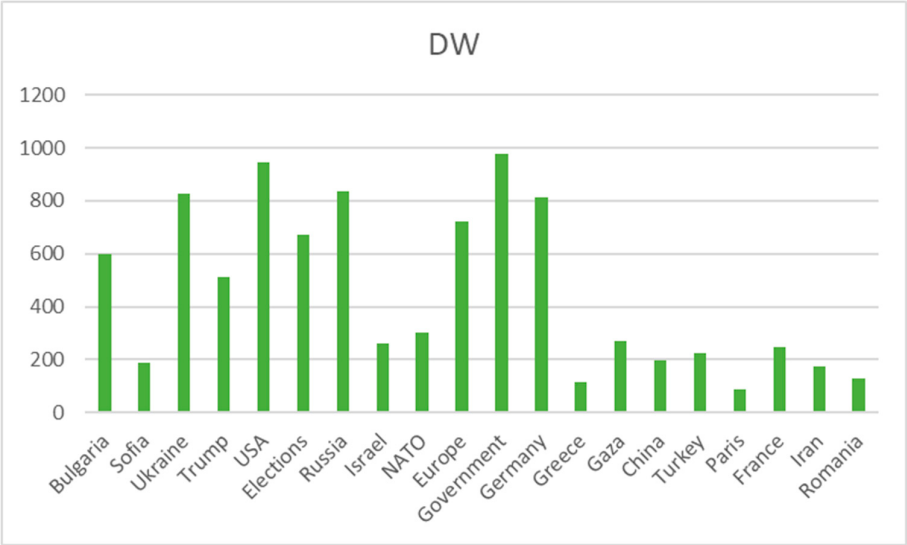


Figure 21. Graphical Representation of the Distribution of Top 20 Words in Publications of DW (April 2024–March 2025).

The data for DW – Bulgarian service, as an international broadcaster, reflect a clearly global editorial focus, consistent with Deutsche Welle’s mission to present international perspectives to regional audiences. Domestic keywords such as “Bulgaria” (600) and “Sofia” (187) appear with relatively low frequency, confirming that DW does not prioritize local news reporting. Instead, the editorial emphasis is on broader political, global, and democratic developments. Terms such as “Government” (978), “USA” (946), “Ukraine” (827), “Russia” (835), and “Germany” (815) are among the most frequent, suggesting sustained coverage of international political relations, war and conflict, governance, and EU affairs. Germany’s presence, not surprisingly, is notable and likely influenced by DW’s country of origin and its lens on EU leadership and diplomacy. “Europe” (722) and “Elections” (674) reinforce DW’s attention to institutional and political processes across the continent. NATO, Israel, Gaza, and China also appear regularly, indicating coverage of geopolitical flashpoints and international security. Compared to other platforms, DW places less emphasis on regional neighbors like Greece, Romania, and Turkey, likely due to its global orientation rather than Balkan-centric focus. In conclusion, DW’s Bulgarian service demonstrates a predominantly international content strategy, reporting on major geopolitical and governance topics while providing minimal coverage of Bulgarian domestic affairs – aligning fully with its role as a global public broadcaster.

Table 20 presents the distribution of top 20 words in publications of Eurocom from April 2024 to March 2025.

Table 20. Distribution of Top 20 Words in Publications of Eurocom (April 2024–March 2025).

Word	Publications
Bulgaria	6251
Sofia	3417
Ukraine	3528
Trump	2446
USA	3736
Elections	3482
Russia	2990
Israel	1430
NATO	1027
Europe	2206
Government	3865

Germany	1013
Greece	570
Gaza	647
China	1041
Turkey	922
Paris	495
France	844
Iran	609
Romania	605

Figure 22 provides a graphical representation of the data shown in Table 20.

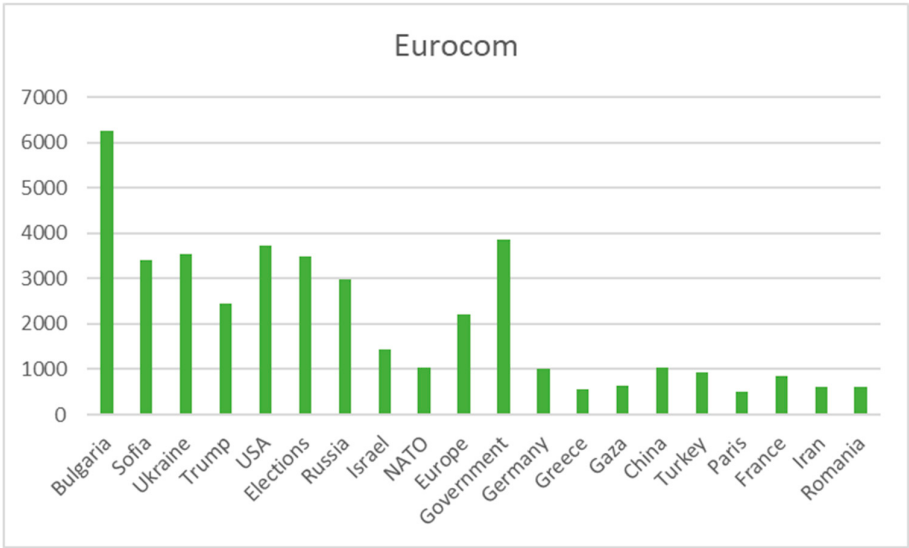


Figure 22. Graphical Representation of the Distribution of Top 20 Words in Publications of Eurocom (April 2024–March 2025).

The lexical profile of Eurocom, a private Bulgarian media outlet with a distinct editorial voice, shows a strong engagement with national political content, paired with considerable attention to international developments. The high number of references to “Bulgaria” (6,251), “Sofia” (3,417), “Government” (3,865), and “Elections” (3,482) clearly reflects a dominant focus on domestic politics, governance, and civic affairs. This positions Eurocom as an outlet heavily invested in political commentary and national discourse. However, Eurocom also dedicates substantial coverage to international affairs, with terms like “USA” (3,736), “Ukraine” (3,528), “Russia” (2,990), and “Trump” (2,446) appearing at high frequencies. This mix suggests an editorial strategy that combines internal political analysis with global context – particularly the international dimensions of security, diplomacy, and foreign influence. The presence of “NATO” (1,027) and “Europe” (2,206) reinforces Eurocom’s attention to Bulgaria’s place within regional and alliance structures, while Middle Eastern topics like “Israel”, “Gaza”, and “Iran” are present but not dominant. Mentions of “China” (1,041) and “Germany” (1,013) reflect interest in key global powers, while references to neighboring countries like “Turkey”, “Greece”, and “Romania” support a well-rounded editorial scope that includes regional relevance. In summary, Eurocom demonstrates a dual editorial profile: deeply focused on domestic political life, while simultaneously maintaining a meaningful presence in international and geopolitical reporting. This combination positions the outlet as both nationally engaged and globally aware.

Table 21 presents the distribution of top 20 words in publications of Nova from April 2024 to March 2025.

Table 21. Distribution of Top 20 Words in Publications of Nova (April 2024–March 2025).

Word	Publications
Bulgaria	9746
Sofia	5054
Ukraine	2544
Trump	1925
USA	3435
Elections	3901
Russia	2015
Israel	1074
NATO	815
Europe	2587
Government	3864
Germany	1319
Greece	879
Gaza	751
China	686
Turkey	900
Paris	906
France	1113
Iran	553
Romania	824

Figure 23 provides a graphical representation of the data shown in Table 21.

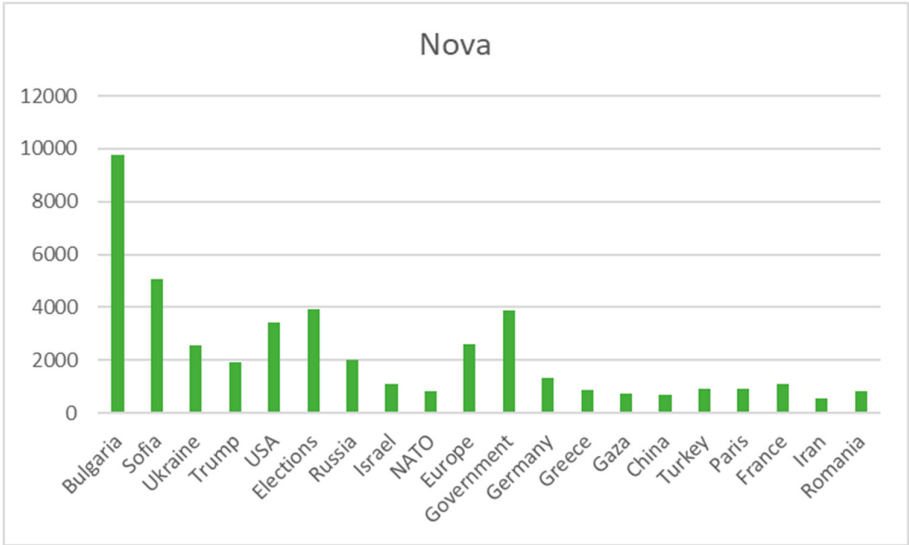


Figure 23. Graphical Representation of the Distribution of Top 20 Words in Publications of Nova (April 2024–March 2025).

The lexical profile of Nova TV, one of Bulgaria’s major commercial broadcasters, reflects a content strategy that strongly integrates national political focus with timely international coverage. Terms like “Bulgaria” (9,746), “Sofia” (5,054), “Government” (3,864), and “Elections” (3,901) show that Nova maintains high engagement with domestic politics and governance, reflecting the editorial weight placed on informing viewers about internal developments and institutional processes. Internationally, Nova provides moderate but consistent coverage of global topics. Frequent mentions of “USA” (3,435), “Ukraine” (2,544), “Russia” (2,015), and “Europe” (2,587) suggest a broad

awareness of key geopolitical developments, especially those with regional and security implications. Nova also demonstrates engagement with event-driven topics, such as “Trump” (1,925) and “Paris” (906) – the latter likely driven by the 2024 Summer Olympics. Similarly, “France” (1,113) and “Germany” (1,319) show up with relatively high frequency, reflecting the importance of EU-centric coverage. Mentions of “Israel”, “Gaza”, “Iran”, and “NATO” confirm that Nova addresses international conflicts and alliances, though not with the intensity seen in more globally focused outlets like Bloomberg or DW. The presence of regional neighbors – “Greece”, “Turkey”, and “Romania” – further supports Nova’s commitment to reporting with both national and regional relevance. In conclusion, Nova TV offers a broad editorial mix, balancing domestic political reporting with responsive coverage of global and regional developments. Its content reflects a mainstream media profile, tailored to a general audience with interest in both local and international affairs.

The results obtained through quantitative and lexical analysis of web publications from eight major Bulgarian media platforms over the period April 2024 – March 2025 allow for several key experimental conclusions that highlight both overarching patterns and distinctive editorial strategies. The analysis of publication volume showed significant variation among platforms, with BNR, Nova, and BNT emerging as the most prolific sources of digital content. Seasonal patterns were evident, particularly with reduced output during the summer months and increased activity in the autumn and early winter, coinciding with periods of heightened political and social developments. The examination of the most frequently used words in titles and full texts revealed a consistent editorial focus on both domestic and international topics. High-frequency terms such as “Bulgaria”, “Sofia”, and “Government” underscore the media’s engagement with national affairs, while terms like “USA”, “Ukraine”, “Russia”, and “Europe” reflect the salience of global geopolitical developments in Bulgarian media discourse. Comparative analysis between word usage in titles and in article bodies revealed editorial strategies designed to capture audience attention. Certain words, including “Trump” and “NATO”, appeared more frequently in titles than in article content, likely chosen for their attention-grabbing quality. In contrast, words such as “Government” and “Bulgaria” were more dominant in the texts themselves, suggesting that these topics receive more substantive treatment within the body of the articles. The monthly dynamics of the top ten keywords illustrated clear, event-driven fluctuations. For instance, coverage of “Trump” and “USA” peaked during the U.S. election period, while the terms “Elections” and “Government” became more prominent during domestic political events. Other terms, such as “Paris”, spiked in connection with specific international occasions like the 2024 Summer Olympics. Finally, the distribution of keywords across individual media platforms demonstrated clear editorial distinctions. Public service broadcasters such as BNR and BNT exhibited a strong focus on domestic issues and political coverage. In contrast, internationally oriented media such as Bloomberg, DW, and Bulgaria ON AIR emphasized global political and economic developments. Private national outlets including bTV, Nova, and Eurocom adopted a more balanced approach, combining consistent local reporting with selective but substantial international coverage.

4. Discussion

The results of this study offer valuable insights into the thematic structure, editorial focus, and content distribution patterns of Bulgarian digital media. Through the application of media intelligence methodologies – including web scraping, NLP-based lexical analysis, and longitudinal observation – the research illustrates how digital news platforms respond to both national developments and global events. The findings support the initial working hypothesis that editorial priorities across media outlets are shaped by a combination of institutional role (public vs. private), audience orientation, and geopolitical context. The observed variation in publication volume between media platforms aligns with prior studies highlighting structural differences between public service broadcasters and commercially driven outlets [4,5]. BNR and BNT, for instance, prioritize coverage of domestic issues, which is reflected in the frequency of terms such as “Bulgaria”, “Sofia”, and “Government”. These patterns confirm their public service mandate and reinforce previous findings

on the content strategies of national broadcasters. In contrast, Bloomberg and DW exhibit stronger international orientation, placing editorial emphasis on global actors and institutions – such as the “USA”, “China”, “Russia”, and “Europe” – which corresponds with their positioning as international or financially focused media. The divergence in word frequency between titles and body texts supports another hypothesis regarding editorial framing. The elevated presence of attention-grabbing words like “Trump” and “NATO” in headlines suggests a strategic approach aimed at increasing visibility and engagement, consistent with findings from recent studies on headline construction and audience targeting [6,9]. On the other hand, more contextually rich and policy-relevant terms, such as “Government”, dominate the article content, pointing to substantive coverage beyond click-oriented framing. The temporal analysis of the top keywords further reinforces the event-driven nature of news production. Fluctuations in terms such as “Elections”, “Ukraine”, and “USA” clearly correspond with political cycles, international crises, and major diplomatic developments. This observation is consistent with earlier work on media responsiveness to political agendas and global events [10,13]. Beyond confirming existing models of media behavior, the study also reveals important nuances in the Bulgarian context. Despite the global trends of digital acceleration and content homogenization, local editorial differences remain strong, shaped by institutional, political, and cultural factors. Although this research does not focus on sentiment classification, it shares common ground with computational approaches that extract meaning from unstructured text data. Similar frameworks have been used in commercial contexts to evaluate large-scale customer feedback [25], demonstrating the generalizability of full-text analysis across domains. This suggests that even in a digitally networked environment, national media systems retain distinct identities and strategic orientations. In a broader context, these findings contribute to the understanding of how digital media ecosystems operate in small states within the European Union. They also highlight the relevance of media intelligence tools for systematically capturing and interpreting large volumes of unstructured news data. Future research could expand this analysis by incorporating multimodal content (video transcripts, social media engagement metrics), applying sentiment analysis and clustering models at the paragraph or sentence level, or comparing results with media landscapes in other countries. Moreover, investigating the role of algorithms and audience data in shaping editorial decisions would provide further depth to the study of media dynamics in the digital age.

5. Conclusions

This study applied a media intelligence framework to analyze the thematic structure and editorial dynamics of Bulgarian digital media through data-driven methods, including web scraping, quantitative content analysis, and natural language processing. The findings demonstrate the complex interplay between domestic and international news coverage, editorial positioning, and media typology. The analysis revealed clear distinctions between public service broadcasters, international outlets, and private national media, each exhibiting unique content patterns and thematic priorities. While public media emphasized national politics and governance, international platforms focused more heavily on geopolitical events and global actors. Private media maintained a hybrid strategy, balancing domestic relevance with responsiveness to major international developments. The results also highlight the significance of title construction in shaping reader attention and reflect the event-driven nature of editorial focus, particularly in the context of political and geopolitical cycles. These insights contribute to a deeper understanding of content strategies in the digital news ecosystem and demonstrate the utility of media intelligence approaches for systematic media monitoring and analysis. From a methodological perspective, the study illustrates the power of web scraping as a scalable and efficient technique for collecting large volumes of real-time media content directly from public news websites. This approach enables timely and flexible data acquisition, bypassing limitations of closed or incomplete archival systems, and offers strong potential for replicability and cross-country comparisons. The findings have broader implications for media professionals, researchers, and policymakers interested in news production, media pluralism,

and digital communication trends. By capturing variations in editorial focus and content dynamics across platforms, the study offers a valuable foundation for future research on agenda-setting, framing strategies, and the intersection between media and political environments – particularly within smaller EU member states and post-transition democracies. This research underscores the critical role of data-driven approaches in understanding how digital media systems function and evolve, while also offering practical tools and insights for navigating the increasingly complex information environment.

Funding: This work was financially supported by the UNWE Research Programme (Research Grant No. 22/2024/A).

Data Availability Statement: The datasets analyzed during this study are available upon request from the author.

Conflicts of Interest: The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

BNR	Bulgarian national radio
BNT	Bulgarian national television
DW	Deutsche Welle (Bulgarian edition)

References

1. McCombs, M.; Shaw, D. L. The Agenda-Setting Function of Mass Media. *Public Opin. Q.* **1972**, *36*(2), 176–187. <https://doi.org/10.1086/267990>.
2. Hallin, D. C.; Mancini, P. *Comparing Media Systems: Three Models of Media and Politics*; Cambridge University Press: Cambridge, 2004.
3. Open Society Institute. *Mapping Digital Media: Bulgaria*; Open Society Foundations, 2013. <https://www.opensocietyfoundations.org/publications/mapping-digital-media-bulgaria>.
4. Cordeiro, D.; Lopezosa, C.; Guallar, J. A Methodological Framework for AI-Driven Textual Data Analysis in Digital Media. *Future Internet* **2025**, *17*, 59. <https://doi.org/10.3390/fi17020059>
5. Tanasescu, L.G.; Vines, A.; Bologa, A.R.; Vaida, C.A. Big Data ETL Process and Its Impact on Text Mining Analysis for Employees' Reviews. *Appl. Sci.* **2022**, *12*, 7509. <https://doi.org/10.3390/app12157509>
6. Kim, E.-G.; Chun, S.-H. Analyzing Online Car Reviews Using Text Mining. *Sustainability* **2019**, *11*, 1611. <https://doi.org/10.3390/su11061611>
7. Hassanien, H.E.-D. Web Scraping Scientific Repositories for Augmented Relevant Literature Search Using CRISP-DM. *Appl. Syst. Innov.* **2019**, *2*, 37. <https://doi.org/10.3390/asi2040037>
8. Naing, I.; Aung, S.T.; Wai, K.H.; Funabiki, N. A Reference Paper Collection System Using Web Scraping. *Electronics* **2024**, *13*, 2700. <https://doi.org/10.3390/electronics13142700>
9. Zherlitsyn, D.; Kolarov, K.; Reкова, N. Digital Transformation in the EU: Bibliometric Analysis and Digital Economy Trends Highlights. *Digital* **2025**, *5*, 1. <https://doi.org/10.3390/digital5010001>
10. Barbera, G.; Araujo, L.; Fernandes, S. The Value of Web Data Scraping: An Application to TripAdvisor. *Big Data Cogn. Comput.* **2023**, *7*, 121. <https://doi.org/10.3390/bdcc7030121>
11. Sarker, K.U.; Saqib, M.; Hasan, R.; Mahmood, S.; Hussain, S.; Abbas, A.; Deraman, A. A Ranking Learning Model by K-Means Clustering Technique for Web Scraped Movie Data. *Computers* **2022**, *11*, 158. <https://doi.org/10.3390/computers11110158>
12. Alexandrescu, A.; Butincu, C.N. Decentralized News-Retrieval Architecture Using Blockchain Technology. *Mathematics* **2023**, *11*, 4542. <https://doi.org/10.3390/math11214542>
13. Haarman, T.; Zijlema, B.; Wiering, M. Unsupervised Keyphrase Extraction for Web Pages. *Multimodal Technol. Interact.* **2019**, *3*, 58. <https://doi.org/10.3390/mti3030058>

14. JavaFX. Available online: <https://openjfx.io/>.
15. jsoup: Java HTML Parser. Available online: <https://jsoup.org/>.
16. Trilling, D.; Jonkman, J. G. F. Scaling Up Content Analysis: How to Manually Annotate a Small Subsample for Supervised Machine Learning. *Soc. Sci. Comput. Rev.* **2018**, *36*(5), 598–614. <https://doi.org/10.1177/0894439318788315>.
17. Bulgaria ON AIR. Available online: <https://bgonair.bg/>.
18. Bloomberg. Available online: <https://bloombergtv.bg/>.
19. BNR News. Available online: <https://bnr.bg/>.
20. BNT News. Available online: <https://bntnews.bg/>.
21. bTV News. Available online: <https://btvnovinite.bg/>.
22. DW. Available online: <https://www.dw.com/bg/>.
23. Eurocom. Available online: <https://eurocom.bg/>.
24. Nova News. Available online: <https://nova.bg/news/>.
25. Ashbaugh, L.; Zhang, Y. A Comparative Study of Sentiment Analysis on Customer Reviews Using Machine Learning and Deep Learning. *Computers* **2024**, *13*, 340. <https://doi.org/10.3390/computers13120340>

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.