

Article

Not peer-reviewed version

---

# From Headlines to Thumbnails: Comparative Analysis of Web Publications in Bulgarian Digital Media and YouTube

---

[Plamen Hristov Milev](#)<sup>\*</sup> and Yavor Nikolov Tabov

Posted Date: 18 August 2025

doi: 10.20944/preprints202508.1187.v1

Keywords: digital media; web publications; YouTube; data intelligence; natural language processing; 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.

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.

Article

# From Headlines to Thumbnails: Comparative Analysis of Web Publications in Bulgarian Digital Media and YouTube

Plamen Hristov Milev <sup>1,\*</sup> and Yavor Nikolov Tabov <sup>2</sup>

<sup>1</sup> Department of Information Technologies and Communications, University of National and World Economy, 1700 Sofia, Bulgaria

<sup>2</sup> Department of Information Technologies and Communications, University of National and World Economy, 1700 Sofia, Bulgaria

\* Correspondence: pmilev@unwe.bg; Tel.: +35928195312

## Abstract

In the contemporary digital landscape, news organizations utilize multiple platforms to reach diverse audiences. This study addresses the question of content differentiation by investigating the cross-platform strategies of three leading Bulgarian news agencies to determine if their thematic priorities are consistent or platform-specific. To achieve this, we conducted a quantitative text analysis of headlines from their websites and official YouTube channels using the TF-IDF algorithm. This was supplemented with a qualitative analysis of YouTube thumbnails to assess their strategic visual contribution. The findings reveal a significant strategic divergence. YouTube channels are primarily dedicated to high-impact domestic political news centered on key public figures. In contrast, their official websites feature a much broader thematic scope, covering international conflicts, extensive cultural and regional events, or a mix of politics and economics. The thumbnail analysis further shows they function as a critical visual layer on YouTube, adding emotional context and explicit cues that are not present in text headlines. This research concludes that news agencies do not simply mirror content but strategically adapt it to leverage the unique characteristics and audience expectations of each platform, employing distinct models for their YouTube and web presences.

**Keywords:** digital media; web publications; YouTube; data intelligence; natural language processing; web scraping

---

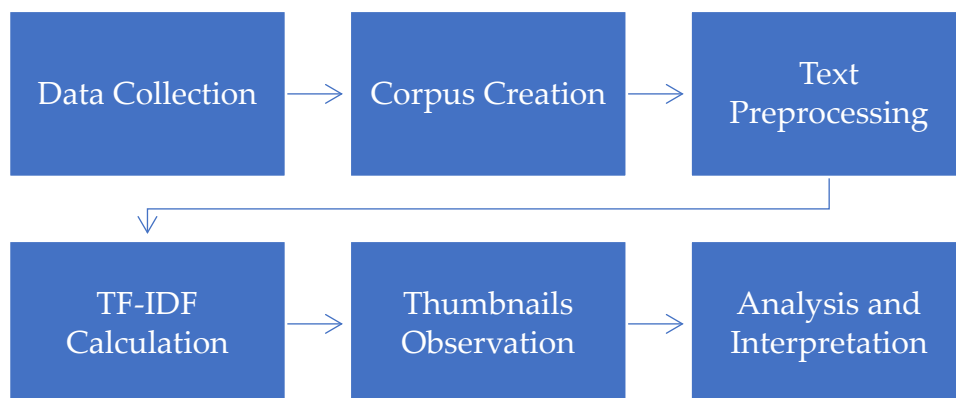
## 1. Introduction

In the contemporary digital media landscape, audiences increasingly rely on visual cues to navigate vast amounts of information [1-6]. Headlines, traditionally the primary method of attracting reader attention, now compete directly with visual elements like thumbnails – small, representative images accompanying digital content. This shift is especially pronounced on platforms such as YouTube, where thumbnails significantly influence user engagement and content discoverability [3-5]. Meanwhile, traditional web-based digital media outlets continue to prioritize textual headlines, albeit increasingly integrating visual strategies inspired by social media platforms. Recent studies have demonstrated the potential of web scraping techniques for collecting and analyzing large-scale data from digital environments, including media and agriculture [1-2]. Research in the domain of digital education and communication highlights the importance of platforms like YouTube, where visual cues are critical in shaping user engagement patterns [6-7]. Moreover, the role of recommender systems on platforms such as YouTube has been analyzed in depth, indicating how algorithmic suggestions contribute to shaping content consumption trends [8]. These developments underline the growing convergence between technological innovations and digital media strategies, as emphasized in recent works exploring the impact of digitalization in higher education and research

[9]. In addition, studies on social media engagement and content dissemination stress the significance of integrating visual components to optimize user attention and interactions [10-11]. The relevance of platforms like YouTube in public opinion formation and information dissemination has been further documented through analyses of comment sections and engagement metrics [12]. In this context, the evolving role of visual content not only reflects changing audience behaviors but also prompts media practitioners to adapt their content strategies accordingly [13-14]. The proliferation of digital platforms has compelled news organizations to move beyond a single-channel distribution model. Platforms like YouTube have become significant news sources, operating in parallel with traditional news websites. This raises a critical question: do news agencies present a uniform editorial voice across platforms, or do they tailor their content to the specific affordances and user behaviors of each channel? This paper addresses this question through a comparative case study of three prominent Bulgarian news agencies with distinct profiles: Bulgarian News Agency (BTA) – also the national news agency, BGNES Agency (a private agency) and Blitz News (a popular tabloid-style agency). We analyze two distinct data corpora: a collection of article headlines from their primary websites and a collection of video titles from their official YouTube channels. The analysis is grounded in data collected through systematic web scraping methods. The objective of the paper is to quantitatively determine the thematic focus of each agency on each platform and to identify patterns of convergence or divergence in their content strategies. We hypothesize that agencies do not simply replicate content but strategically differentiate it, using YouTube for more personality-driven and high-impact news and their websites for a broader, more comprehensive coverage. The primary contributions of this study include providing insights into evolving media presentation practices, establishing a comparative framework for analyzing visual versus textual dominance. Our findings will contribute to a deeper understanding of the intersection between traditional and new media dynamics in the digital era.

## 2. Materials and Methods

This study employs a mixed-method approach to provide a comprehensive analysis of the digital content produced by the three key Bulgarian news agencies, which were selected to represent a diverse spectrum of media models (public, private, and sensationalist). The research process began with data collection via web scraping to systematically gather headlines from official websites and YouTube channels of the agencies. Following this, a quantitative text analysis using the TF-IDF algorithm was applied to identify and rank the most significant keywords, thereby revealing the thematic priorities on each platform. This quantitative stage was complemented by a qualitative assessment of YouTube thumbnails to analyze their strategic function as a layer of visual communication. This integrated methodology ensures a robust, data-driven comparison of the editorial and visual strategies the agencies deploy across different digital platforms. Figure 1 illustrates the schematic of the data analysis pipeline.



**Figure 1.** Data Processing and Analysis Workflow.

To ensure the consistency and validity of the analysis, we applied a selection criterion focused on informational agencies that actively operate across both digital ecosystems – web and YouTube. Specifically, we included only those agencies that maintain regularly updated websites with textual news content and operate official YouTube channels with frequent video publications. YouTube was selected as the primary video-sharing platform for this study due to its dominant position in the global and Bulgarian digital media landscape. As the world’s largest video-hosting service, YouTube offers unparalleled reach, user engagement, and content variety, including a significant presence of professional news organizations. Its standardized metadata structure, advanced search and filtering capabilities, and publicly accessible APIs facilitate consistent data collection and analysis across channels. In Bulgaria, major information agencies maintain active and regularly updated YouTube channels, ensuring both the availability of high-quality audiovisual content and a direct link to their official editorial policies. These characteristics make YouTube uniquely suitable for conducting a systematic and comparable cross-platform analysis between video-based and text-based news content. After a preliminary review of the Bulgarian media landscape, three informational agencies met the requirements: Bulgarian News Agency, BGNES Agency and Blitz News. These agencies were chosen because they enable a direct comparison between web publications (headlines, article structure, textual features) and YouTube content (titles, thumbnails, visual and auditory cues) under a unified editorial and organizational framework. The Bulgarian News Agency is the national public information agency of Bulgaria, providing comprehensive news coverage across politics, economy, culture, and international affairs. The official BTA website delivers up-to-date news from Bulgaria and around the world, focusing on socially significant events, economics, politics, and culture [15]. The official BTA YouTube channel features video reports, press conferences, and interviews, often complementing the website’s content with multimedia formats [16]. BGNES Agency is an independent news agency, offering timely reports and analyses on domestic and global developments. The BGNES website offers national and international news, analyses, and reports, with an emphasis on balanced coverage of current events [17]. The official BGNES YouTube channel includes video interviews, reports, and press events, synchronized with the publications on the website [18]. Blitz News is a privately owned Bulgarian news agency, delivering breaking news, exclusive interviews, and multimedia content with a focus on national and international events. The Blitz website publishes high-frequency news covering a wide range of topics – from politics and economics to lifestyle and sports [19]. The Blitz YouTube channel presents video materials, news bulletins, and reports that visually enrich and complement the articles published on the website [20]. The three agencies represent distinct types of media ownership and editorial orientation, namely public, private, and commercially driven, respectively, which allows for a richer, more nuanced analysis of cross-platform content strategies. The selected approach ensures that the study remains focused on comparable cases, thereby avoiding asymmetries in data representation or platform activity. Table 1 presents the profile of news agencies selected for research.

**Table 1.** Characteristics of Selected Media Sources.

Media	Type	Style/tone
Bulgarian News Agency	Public	Formal, factual, journalistic
BGNES Agency	Private	Aggressive, commercial, media flow
Blitz News	Private	Sensational, clickbait, visually intense

All data used in this study were collected from publicly accessible sources and processed exclusively for academic, non-commercial research purposes. The analysis includes only metadata such as headlines, titles, thumbnails (descriptive analysis only), and basic publication metrics (e.g., publication date), without reproducing full textual content or audiovisual material. In accordance with copyright laws and fair use principles applicable to scholarly research, the study does not make

use of protected media content. Article headlines and video titles were used solely for analytical purposes. This approach ensures compliance with intellectual property rights while enabling comparative analysis of cross-platform communication strategies. Initially, the data was collected through systematic web scraping methods to collect relevant data [21-22]. The data set comprises content published over a one-year period from July 2024 to June 2025. The data collection focused primarily on capturing textual headlines, associated thumbnails and publication dates. Similar methodological approaches for data extraction from online sources have been described in previous studies focused on digital education platforms and open-source data collection [23-24]. In the case of YouTube, data was obtained firstly through web scraping for identifying the video publications and subsequently using YouTube Data API. The use of API-based data extraction and web scraping methods has been extensively discussed in the context of collecting user-generated content for media analysis and sentiment studies [25-26]. Ethical considerations regarding data collection and analysis were strictly observed, ensuring compliance with terms of service of the respective platforms, and no personally identifiable information was collected or processed. Similar adherence to ethical standards in digital data research has been highlighted in recent publications on media and sentiment analysis [27-28]. Additional research has demonstrated the applicability of web scraping techniques in the field of applied sciences, particularly for the extraction and preprocessing of large-scale datasets [29]. Such approaches have been effectively used to support decision-making processes through systematic data collection and analysis in related studies [30]. Furthermore, the potential of automated methods for extracting structured data from online environments is well established in recent literature on applied systems innovation [31]. The data for this study were collected through web scraping between 1<sup>st</sup> of July 2024 and 30<sup>th</sup> of June 2025. A custom-built data-gathering application, developed in the Java programming language, was engineered for this task. The application was designed to systematically send HTTP requests to the URLs of the selected news websites and YouTube channels. Upon receiving the server response, the application parsed the raw HTML content to locate and extract the required data points, specifically the headlines and video titles. To ensure ethical scraping practices and avoid server overload, a rate limit was implemented by introducing a deliberate delay between consecutive requests. The extracted raw data were then cleaned and structured into CSV files, creating the two distinct corpora for analysis (websites and YouTube). Each corpus represents the complete collection of all text documents (in our case, all post titles) that will be analyzed. To prepare the data for analysis, each title in the corpus went through the several processing steps. All letters were converted to lowercase to ensure that words such as "Government" and "government" were treated as the same term. Each title was divided (tokenized) into individual words called "tokens". All punctuation marks and numbers were removed, as they did not carry any semantic value for the purposes of this analysis. A filter was applied to remove common words in the Bulgarian language that do not carry any specific meaning. This allows the analysis to focus on the words that actually define the topic of the text. The core of the analysis is the calculation of the TF-IDF value for each token in each document. The concept of TF-IDF is fundamental in the field of Information Retrieval and Natural Language Processing (NLP). The definition of the concept of Inverse Document Frequency argues that the specificity of a term (and therefore its weight) is inversely proportional to the frequency with which it occurs in the documents of a given collection [32]. This concept has a direct relationship to the Vector Space Model, in which documents are represented as vectors in a multidimensional space. TF-IDF is the most used method for calculating the weights of the components of these vectors [33]. TF-IDF and its variations have significant applications in search engines and information systems [34]. The value of TF-IDF is the product of two metrics: Term Frequency (TF) and Inverse Document Frequency (IDF). Term Frequency measures how often a given word appears in a particular document (title). It is normalized by dividing the number of occurrences by the total number of words in the document, so as not to favor longer documents:

$$TF(t, d) = \text{Number of times term } t \text{ appears in document } d / \text{Total number of terms in document } d \quad (1)$$

Inverse Document Frequency measures the importance of a word in the entire corpus. It gives higher weight to words that occur in few documents and lower weight to those that occur frequently throughout. It is calculated as the logarithm of the ratio between the total number of documents in the corpus and the number of documents containing the term:

$$IDF(t, D) = \log(\text{Total number of documents in corpus } D / (1 + \text{Number of documents containing term } t)) \quad (2)$$

Adding "+1" to the denominator is a standard "smoothing" practice that prevents division by zero. Conceptually, it ensures that any new terms encountered during analysis are assigned a high, finite IDF score, correctly reflecting their informational novelty and specificity. The final value of TF-IDF for each word in each document is obtained by multiplying its TF and IDF values:

$$TF-IDF(t, d, D) = TF(t, d) * IDF(t, D) \quad (3)$$

After the TF-IDF matrix (containing the values for each word in each document) was calculated, we proceeded to aggregate the results. To identify the top keywords for each agency and platform, the mean TF-IDF score for each term was calculated across all relevant documents. Terms with the highest average scores were ranked as the most significant. For the YouTube dataset, a supplementary analysis was conducted on the video thumbnails associated with top-ranking headlines to assess their strategic role in adding context and emotional value.

### 3. Results

The findings of our research are organized into two main parts. First, the results of the quantitative keyword analysis, derived from the TF-IDF algorithm, detail the thematic priorities on each platform. Second, a qualitative analysis of the added value of YouTube thumbnails examines their strategic role in visual communication.

Table 2 presents the total number of publications for each of the media platforms analyzed within the period from July 2024 to June 2025.

**Table 2.** Total Number of Publications by Agency (July 2024–June 2025).

Digital Media	Website	Publications
Bulgarian News Agency	<a href="https://bta.bg">https://bta.bg</a>	206,950
BGNES Agency	<a href="https://bgnes.bg">https://bgnes.bg</a>	21,492
Blitz News	<a href="https://blitz.bg/">https://blitz.bg/</a>	86,661

A striking observation is the massive output of the Bulgarian News Agency, which published 206,950 articles during the specified period. This volume is substantially higher than that of the private agencies, being nearly ten times greater than BGNES Agency output and more than double that of Blitz News. This aligns with BTA's mandate as a national news agency, which is tasked with providing comprehensive, non-selective coverage of a wide range of events. In contrast, Blitz and BGNES, while still prolific, demonstrate a more editorially focused output, which is characteristic of private media organizations that prioritize specific topics.

Table 3 presents the total number of videos for each of the analyzed YouTube channels within the period from July 2024 to June 2025.

**Table 3.** Total Number of Videos by Agency (July 2024–June 2025).

Channel	Homepage	Videos
Bulgarian News Agency	<a href="https://www.youtube.com/@BulgarianNewsAgency">https://www.youtube.com/@BulgarianNewsAgency</a>	5285

BGNES Agency	<a href="https://www.youtube.com/@BGNESAAgency">https://www.youtube.com/@BGNESAAgency</a>	1095
Blitz News	<a href="https://www.youtube.com/@BlitzBGNews">https://www.youtube.com/@BlitzBGNews</a>	581

The data illustrates a different strategic landscape that emerges. While BTA maintains its position as the most active content creator with 5,285 videos, there is a notable reversal in the ranking of the two private agencies. Blitz, despite publishing four times as many web articles as BGNES, produced significantly fewer videos (581 compared to 1,095 by BGNES). This disparity suggests that BGNES has a more developed or prioritized video production strategy for its YouTube channel relative to its overall content output. Conversely, Blitz's content strategy appears to be heavily concentrated on its text-based website, with video playing a less central role. These figures indicate that the agencies' resource allocation and strategic priorities differ significantly between their web and video platforms.

### 3.1. Keyword Analysis (TF-IDF)

The TF-IDF analysis yielded distinct keyword hierarchies for each platform. The scores represent the mean TF-IDF value for each term, indicating its overall importance within the respective corpus.

Table 4 presents the top 15 ranked words in publication titles of the agencies' websites, during the analyzed period (July 2024-June 2025), sorted by their score in descendent order.

**Table 4.** Top 15 Keywords in Titles of Agencies' Websites (July 2024–June 2025).

Keyword	Mean TF-IDF Score
ukrayna (Ukraine)	0.038
borisov	0.035
voynata (the war)	0.026
rusiya (Russia)	0.025
radev	0.023
pravitelstvo (government)	0.023
kiril	0.022
petkov	0.022
sofia	0.021
tsena (price)	0.021
ukrainski (Ukrainian)	0.019
ruski (Russian)	0.018
gaz (gas)	0.018
putin	0.017
evropeyskiya (the European)	0.016

The results clearly show that there is a very strong presence of the international topic, specifically the war in Ukraine ("Ukraine", "war", "Russia", "Putin"). Domestic politics remains important but shares the leading place.

Table 5 presents the top 10 ranked words in publication titles of Bulgarian News Agency on their website, during the analyzed period (July 2024-June 2025), sorted by their score in descendent order.

**Table 5.** Top 10 Keywords in Titles of Bulgarian News Agency (July 2024–June 2025).

Keyword	Mean TF-IDF Score
teatur (theatre)	0.045
izlozhba (exhibition)	0.044
kontsert (concert)	0.038
varna	0.031
evropeyskiya (the European)	0.028
bulgaria	0.027

obshtina (municipality)	0.027
mezhdunarodniya (the international)	0.026
festival (fest)	0.025
plovdiv	0.024

BTA emphasizes cultural events ("theater", "exhibition", "concert"), regional news ("Varna", "Plovdiv") and official EU topics ("European"). This corresponds to its function as a national agency with broad coverage.

Table 6 presents the top 10 ranked words in publication titles of BGNES Agency on their website, during the analyzed period (July 2024-June 2025), sorted by their score in descendent order.

**Table 6.** Top 10 Keywords in Titles of BGNES Agency (July 2024–June 2025).

Keyword	Mean TF-IDF Score
borisov	0.052
tsena (price)	0.046
gaz (gas)	0.042
radev	0.040
pravitelstvo (government)	0.039
kiril	0.038
petkov	0.038
gerb (political party)	0.035
asen	0.034
vasilev	0.034

BGNES Agency focuses almost entirely on domestic politics and economics ("Borisov", "price", "gas", "Radev", "government").

Table 7 presents the top 10 ranked words in publication titles of Blitz News on their website, during the analyzed period (July 2024-June 2025), sorted by their score in descendent order.

**Table 7.** Top 10 Keywords in Titles of Blitz News (July 2024–June 2025).

Keyword	Mean TF-IDF Score
ukrayna (Ukraine)	0.051
voynata (the war)	0.043
rusiya (Russia)	0.039
borisov	0.038
putin	0.032
shok (shock)	0.028
tragediya (tragedy)	0.027
kiev	0.027
ruskata (the Russian)	0.025
petkov	0.024

Blitz News focuses on the war in Ukraine, using strong and emotional words ("shock", "tragedy"). Domestic politics are secondary.

Table 8 presents the top 15 ranked words in video titles of the agencies' YouTube channels, during the analyzed period (July 2024-June 2025), sorted by their score in descendent order.

**Table 8.** Top 15 Keywords in Titles of Agencies' YouTube Channels (July 2024–June 2025).

Keyword	Mean TF-IDF Score
borisov	0.053
pravitelstvo (government)	0.040

kiril	0.038
petkov	0.038
radev	0.036
izbori (elections)	0.033
rumen	0.032
gerb (political party)	0.031
asen	0.031
vasilev	0.031
boyko	0.029
mandata (the mandate)	0.026
glasuva (votes)	0.024
proektokabineta (the project-cabinet)	0.023
zhelyazkov	0.022

The results clearly show that the political situation in the country is a dominant topic. The names of key political figures (Borisov, Petkov, Radev, Vassilev), parties (GERB) and processes (government, elections, mandate) occupy the top spots.

Table 9 presents the top 10 ranked words in video titles of Bulgarian News Agency on their YouTube channel, during the analyzed period (July 2024-June 2025), sorted by their score in descendent order.

**Table 9.** Top 10 Keywords in Titles of Bulgarian News Agency on YouTube (July 2024–June 2025).

Keyword	Mean TF-IDF Score
pravitelstvo (government)	0.095
proektokabineta (the project-cabinet)	0.063
glasuva (votes)	0.063
bulgaria	0.056
zhelyazkov	0.053
radev	0.052
mandata (the mandate)	0.049
rumen	0.046
pravitelstvoto (the government)	0.044
ministri (ministers)	0.040

BTA focuses on the institutional and formal aspects of politics. Words like "government," "draft cabinet," "vote," "mandate," "ministers" indicate a focus on the processes of state governance, rather than on individuals. This is expected by a national news agency. The difference with BTA compared to the publications on their website is huge. Their YouTube channel is strictly focused on official political processes (the word "government" has a very high rating). Their website, on the other hand, has a much broader coverage and serves as a national information source on culture, regions and official European topics. Political news is only part of the general flow.

Table 10 presents the top 10 ranked words in video titles of BGNES Agency on their YouTube channel, during the analyzed period (July 2024-June 2025), sorted by their score in descendent order.

**Table 10.** Top 10 Keywords in Titles of BGNES Agency on YouTube (July 2024–June 2025).

Keyword	Mean TF-IDF Score
borisov	0.075
izbori (elections)	0.066
gurtsiya (Greece)	0.056
pravitelstvo (government)	0.055
gerb (political party)	0.048
boyko	0.046

asen	0.045
vasilev	0.045
evrozonata (the eurozone)	0.042
finansi (finances)	0.042

BGNES Agency occupies an intermediate position. It covers both political figures ("Borisov", "Vassilev") and processes ("elections", "government"). Specific economic and regional topics such as "the eurozone", "finance" and "Greece" also appear, which indicates a broader reporting scope. BGNES Agency is consistent in its content. In both channels (website and YouTube) the focus is on domestic politics and economic topics. The keywords are almost identical, with the website showing a slightly stronger emphasis on specific economic issues such as "price" and "gas", while on YouTube broader topics such as "eurozone" and international relations ("Greece").

Table 11 presents the top 10 ranked words in video titles of Blitz News on their YouTube channel, during the analyzed period (July 2024-June 2025), sorted by their score in descendent order.

**Table 11.** Top 10 Keywords in Titles of Blitz News on YouTube (July 2024–June 2025).

Keyword	Mean TF-IDF Score
borisov	0.076
petkov	0.054
kiril	0.054
cska (football club)	0.052
radev	0.045
asen	0.042
vasilev	0.042
rumen	0.038
boyko	0.037
kostadinov	0.033

Blitz News has a strong focus on political figures and confrontations ("Borisov", "Petkov", "Radev", "Kostadinov"). The presence of sports topics ("CSKA") is also impressive, which distinguishes it from the other two agencies. The language is more direct and personalized. In comparison to their website publications, Blitz's focus is changing dramatically. YouTube is dominated by domestic political figures and conflicts, as well as sports. However, on their site, they focus primarily on the war in Ukraine, using highly emotional language. Domestic politics remains but is given a lower priority.

These quantitative results highlight clear differences in the agencies' thematic focus. YouTube is used primarily to cover dynamic political events and personalized conflicts, probably because the video format is more suitable for this. Websites allow for broader and more diverse coverage. Some (like Blitz News) use it to cover leading international news, while others (like BTA) use it to fully fulfill their role as a national agency with cultural and regional news. BGNES Agency maintains the most similar editorial policy on both platforms. However, the full strategic picture on YouTube is incomplete without considering the visual layer of communication, which is analyzed next.

### 3.2. YouTube Thumbnails Added Value Analysis

The main goal of any thumbnail is to increase the CTR (Click-Through Rate) by getting the user to click on the video. Unlike a standard title, a photo achieves this through emotional impact. Website titles inform, while thumbnails evoke emotion. Most successful thumbnails show the faces of the key figures in the news (in our case, Borisov, Petkov, Radev, etc.). The human brain is designed to respond to faces and emotions. Photos are often selected to show strong emotion – anger, concern, surprise, triumph. The title "Borisov with a comment on the government" is neutral. But a thumbnail with Borisov's angry face instantly creates a sense of conflict and drama, which provokes curiosity. A

photo can convey a huge amount of information in a second. The thumbnail hints at a story. A photo of a burning building or a line of people instantly communicates the topic, even before the user has read the title. This makes the news easier to "digest". A short, bold text in capital letters is often placed above the photo. This text does not repeat the title but complements it with its most provocative part – often a quote or question. By consistently using logos, color schemes and fonts on thumbnails, media builds its visual identity. Users begin to recognize the style of a given media just by the photo, which builds loyalty and faster orientation in the YouTube feed. While websites rely on SEO and factual content, YouTube channels struggle for attention in a highly competitive visual environment. Thumbnails are their most powerful weapon in this fight.

Table 12 outlines the observed visual strategies employed by each news agency on their YouTube channels, linking them to their dominant content themes and the resulting strategic value. The analysis is based on the correlation between the top-ranking keywords (from the TF-IDF results) and the associated thumbnails.

**Table 12.** Strategic Analysis of YouTube Thumbnail Value by News Agency.

Channel	Dominant Keyword Themes	Observed Thumbnail Strategy	Added Value
Bulgarian News Agency	Institutional Processes, Formal Politics	Formal, objective imagery depicting institutional symbols (e.g., the Parliament Hall, official podiums, flags) rather than individual politicians' faces. Clean, professional aesthetic with minimal text overlays.	Authority & Objectivity: Reinforces the agency's brand as a formal, unbiased national news source. Provides context of scale and officiality, appealing to an audience seeking formal information.
BGNES Agency	Domestic & Economic Policy, International Relations	Conceptual, often collage-style images that juxtapose a key person (e.g., a politician) with a visual symbol of the topic (e.g., the Euro symbol, national flags).	Contextualization & Accessibility: Visualizes complex or abstract topics, making them more understandable. Directly links figures to the issues they are discussing, providing clear context.
Blitz News	Political Confrontation, Sports Figures	Emotional, close-up shots of key figures displaying strong emotions (e.g., anger, frustration). Frequent use of large, provocative text overlays (e.g., "LIAR!", "SCANDAL!").	Dramatization & Personalization: Transforms news reports into personal conflicts, maximizing emotional engagement and creating a strong incentive for clicks (high CTR appeal).

The analysis of the headline-thumbnail relationship on the data obtained confirms and concretizes our initial expectations. BTA uses thumbnails to show the scale and officiality of events, emphasizing institutional symbols. BGNES Agency uses thumbnails to connect individuals to specific topics and to visualize complex concepts, often through collages. Blitz News uses thumbnails to dramatize and emotionally amplify the news, focusing on faces and conflicts.

#### 4. Discussion

The findings from this study clearly indicate that the news agencies employ distinct, platform-specific content strategies. The agencies' websites reflect a much wider editorial scope. The dominant shared theme is the war in Ukraine, a topic virtually absent from the top keywords on YouTube. BTA undergoes the most dramatic shift. Its website fulfills a true national agency role, with top keywords related to culture ("theatre," "exhibition"), regional news ("Varna," "Plovdiv"), and official EU topics.

Politics is just one of many content streams. BGNES Agency remains the most thematically consistent agency across both platforms. Its website maintains a strong focus on domestic politics and economics, with keywords like "price" and "gas" ranking highly, demonstrating a sustained specialization. Blitz News heavily prioritizes the conflict in Ukraine, using high-emotion keywords like "war," "shock," and "tragedy," aligning with its tabloid style. On YouTube, all three agencies converge on a narrow set of topics centered on domestic political figures and processes. This highlights an adaptive strategy aimed at aligning with the expectations of audiovisual audiences [24]. The high ranking of names like "Borisov," "Petkov," and "Radev" suggests a personality-driven news agenda. BTA, as the national agency, focuses on the formal, institutional aspects of politics, with keywords like "government," "project-cabinet," and "votes" scoring highest. BGNES Agency balances its coverage between key political figures ("Borisov") and overarching political events ("elections"), while also touching on foreign policy ("Greece") and economic matters ("eurozone"). Blitz News adopts a confrontational angle, focusing on political clashes but also carving out a unique niche with sports content ("CSKA"), likely to attract a specific demographic. The headline data only tells part of the story. Although textual headlines maintain their importance, integrating compelling visual elements offers a clear advantage in capturing user attention and driving interactions [35-36]. The thumbnails associated with YouTube videos represent a critical layer of strategic communication, particularly within video-sharing platforms where visual immediacy strongly influences user behavior [37-38]. Based on a qualitative assessment of the thumbnails linked to the top-ranking keywords in our dataset, we observe several specifics. For videos on governmental procedures, BTA favors thumbnails depicting institutional symbols, such as the parliament hall or official podiums. This reinforces its brand of formal, objective reporting. For complex topics, BGNES Agency often uses collage-style thumbnails that juxtapose a key political figure with a visual symbol of the topic (e.g., the EU flag for a story on the Eurozone). This serves to visualize abstract concepts and make them more accessible. For videos about political conflicts, Blitz News consistently uses thumbnails with close-up images of politicians' faces displaying strong emotions (e.g., anger, frustration), often overlaid with provocative text. This transforms a news report into a personal drama. This visual layer is a key differentiator, designed to maximize emotional engagement and CTR in a competitive visual environment – a function that text-based website headlines do not perform. These findings align with prior studies highlighting the impact of content and metadata, including thumbnails, on platform visibility and algorithmic promotion strategies [39]. In practical terms, the study highlights the need for digital content creators to prioritize visual strategies, especially when operating within platforms characterized by high visual competition. This observation is consistent with insights from environmental monitoring and urban studies that stress the role of visual representation in information dissemination [40-41]. Future research may expand upon these findings by exploring additional platforms, varying content genres, and broader geographical contexts to further illuminate the evolving role of visual content in digital communication [28,42]. Moreover, the integration of spatial data and geoinformation systems with online media content analysis opens new avenues for future interdisciplinary research, as highlighted in recent studies [43]. Beyond its academic contribution, this study offers practical insights for media practitioners, communication specialists, and digital strategists. Understanding that audiences on different platforms respond to distinct types of content allows for more effective resource allocation and tailored communication strategies. The findings demonstrate that a "one-size-fits-all" approach to digital news distribution is suboptimal, and agencies can optimize engagement by developing platform-specific editorial and visual guidelines, as evidenced by the clear strategic differentiation between institutional focus and sensationalist approach.

## 5. Conclusions

This study provides quantitative evidence that Bulgarian news agencies do not pursue a monolithic content strategy but rather adapt their editorial focus to the platform. YouTube is leveraged as a high-impact channel for personality-driven domestic political news, optimized for

engagement through emotionally charged visual thumbnails. In contrast, the official websites fulfill a broader informational mandate, covering a wider range of topics including international affairs, economics, and culture, depending on the agency's specific profile. Private media emerges as the most thematically consistent content provider, while public media shows the greatest strategic divergence, clearly separating its formal political reporting on YouTube from its comprehensive cultural and regional coverage on its website. Sensationalist media pivots its primary focus from domestic politics on YouTube to the international conflict in Ukraine on its website, maintaining its sensationalist tone across both. These findings underscore the growing sophistication of news distribution strategies in the digital age. Future research could expand on this analysis by examining user engagement metrics (likes, comments) to measure audience reception or by conducting a longitudinal study to track the evolution of these strategies over time.

**Author Contributions:** Conceptualization, P.M.; methodology, P.M. and Y.T.; software P.M. and Y.T.; validation, P.M. and Y.T.; formal analysis, P.M. and Y.T.; investigation, P.M. and Y.T.; resources, P.M. and Y.T.; data curation, P.M. and Y.T.; writing—original draft preparation, P.M. and Y.T.; writing—review and editing, P.M.; visualization, P.M. and Y.T.; supervision, P.M.; project administration, P.M.; funding acquisition, P.M. All authors have read and agreed to the published version of the manuscript.

**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 corresponding author.

**Conflicts of Interest:** The authors declare 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:

NLP	Natural language processing
BTA	Bulgarian News (Telegraph) Agency
TF	Term Frequency
IDF	Inverse Document Frequency
CTR	Click-Through Rate
SEO	Search Engine Optimization

## References

1. Santos, F.; Acosta, N. An Approach Based on Web Scraping and Denoising Encoders to Curate Food Security Datasets. *Agriculture* **2023**, *13*, 1015. <https://doi.org/10.3390/agriculture13051015>
2. Rodríguez-Almonacid, D.V.; Ramírez-Gil, J.G.; Higuera, O.L.; Hernández, F.; Díaz-Almanza, E. A Comprehensive Step-by-Step Guide to Using Data Science Tools in the Gestion of Epidemiological and Climatological Data in Rice Production Systems. *Agronomy* **2023**, *13*, 2844. <https://doi.org/10.3390/agronomy13112844>
3. Park, Y.; Shin, Y. Novel Scratch Programming Blocks for Web Scraping. *Electronics* **2022**, *11*, 2584. <https://doi.org/10.3390/electronics11162584>
4. Ahmed, E.; Xue, L.; Sankalp, A.; Kong, H.; Matos, A.; Silenzio, V.; Singh, V.K. Predicting Loneliness through Digital Footprints on Google and YouTube. *Electronics* **2023**, *12*, 4821. <https://doi.org/10.3390/electronics12234821>
5. 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>

6. Li, B.; Kou, X.; Bonk, C.J. Embracing the Disrupted Language Teaching and Learning Field: Analyzing YouTube Content Creation Related to ChatGPT. *Languages* **2023**, *8*, 197. <https://doi.org/10.3390/languages8030197>
7. Naganawa, H.; Hirata, E. Social Media and Logistics: Uncovering Challenges and Solutions Through YouTube Data. *Logistics* **2025**, *9*, 56. <https://doi.org/10.3390/logistics9020056>
8. McGarry, K. Analyzing Social Media Data Using Sentiment Mining and Bigram Analysis for the Recommendation of YouTube Videos. *Information* **2023**, *14*, 408. <https://doi.org/10.3390/info14070408>
9. Kirilov, R. Approaches for building information systems for monitoring the realization of students. *Econ. Alt.* **2021**, *3*, 469-481. <https://doi.org/10.37075/EA.2021.3.09>
10. Silvallana, D.F.; Elias, C.; Catalan-Matamoros, D. Exploring Vaccine Hesitancy in the Philippines: A Content Analysis of Comments on National TV Channel YouTube Videos. *Int. J. Environ. Res. Public Health* **2025**, *22*, 819. <https://doi.org/10.3390/ijerph22060819>
11. Bhagat, K.K.; Mishra, S.; Dixit, A.; Chang, C.-Y. Public Opinions about Online Learning during COVID-19: A Sentiment Analysis Approach. *Sustainability* **2021**, *13*, 3346. <https://doi.org/10.3390/su13063346>
12. Osman, N.S.; Kim, J.-H.; Park, J.-H.; Park, H.-W. Identifying the Impacts of Social Movement Mobilization on YouTube: Social Network Analysis. *Information* **2025**, *16*, 55. <https://doi.org/10.3390/info16010055>
13. Musleh, D.A.; Alkhwaja, I.; Alkhwaja, A.; Alghamdi, M.; Abahussain, H.; Alfawaz, F.; Min-Allah, N.; Abdulqader, M.M. Arabic Sentiment Analysis of YouTube Comments: NLP-Based Machine Learning Approaches for Content Evaluation. *Big Data Cogn. Comput.* **2023**, *7*, 127. <https://doi.org/10.3390/bdcc7030127>
14. Luțan, E.-R.; Bădică, C. Emotion-Based Literature Book Classification Using Online Reviews. *Electronics* **2022**, *11*, 3412. <https://doi.org/10.3390/electronics11203412>
15. Bulgarian News Agency. Available online: <https://bta.bg/> (accessed on 1 July 2025).
16. Bulgarian News Agency – YouTube. Available online: <https://www.youtube.com/@BulgarianNewsAgency> (accessed on 1 July 2025).
17. BGNES Agency. Available online: <https://bgnes.bg/> (accessed on 1 July 2025).
18. BGNES Agency – YouTube. Available online: <https://www.youtube.com/@BGNESAgency> (accessed on 1 July 2025).
19. Blitz News. Available online: <https://blitz.bg/> (accessed on 1 July 2025).
20. Blitz News – YouTube. Available online: <https://www.youtube.com/@BlitzBGNews> (accessed on 1 July 2025).
21. 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>
22. Skoulikaris, C.; Krestenitis, Y. Cloud Data Scraping for the Assessment of Outflows from Dammed Rivers in the EU. A Case Study in South Eastern Europe. *Sustainability* **2020**, *12*, 7926. <https://doi.org/10.3390/su12197926>
23. Hayes, D.R.; Cappa, F.; Cardon, J. A Framework for More Effective Dark Web Marketplace Investigations. *Information* **2018**, *9*, 186. <https://doi.org/10.3390/info9080186>
24. Tsiourlini, M.; Tzafilkou, K.; Karapiperis, D.; Tjortjis, C. Text Analytics on YouTube Comments for Food Products. *Information* **2024**, *15*, 599. <https://doi.org/10.3390/info15100599>
25. Lu, J.C.-C. Creating Effective Educational Videos on YouTube in Higher Education. *Eng. Proc.* **2023**, *38*, 32. <https://doi.org/10.3390/engproc2023038032>
26. Giannakouloupoulos, A.; Pergantis, M.; Lamprogeorgos, A.; Lampoura, S. A Data-Driven Approach to Estimating the Impact of Audiovisual Art Events Through Web Presence. *Information* **2025**, *16*, 88. <https://doi.org/10.3390/info16020088>
27. MacLean, C.; Cavallucci, D. Assessing Fine-Tuned NER Models with Limited Data in French: Automating Detection of New Technologies, Technological Domains, and Startup Names in Renewable Energy. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 1953-1968. <https://doi.org/10.3390/make6030096>

28. Polpanich, O.-u.; Bhatpuria, D.; Santos Santos, T.F.; Krittasudthacheewa, C. Leveraging Multi-Source Data and Digital Technology to Support the Monitoring of Localized Water Changes in the Mekong Region. *Sustainability* **2022**, *14*, 1739. <https://doi.org/10.3390/su14031739>
29. 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>
30. Louro, J.; Fidalgo, F.; Oliveira, Â. Recognition of Food Ingredients—Dataset Analysis. *Appl. Sci.* **2024**, *14*, 5448. <https://doi.org/10.3390/app14135448>
31. Hassanién, 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>
32. Sparck Jones, K. A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *J. Doc.* **1972**, *28*, 11–21.
33. Salton, G.; Wong, A.; Yang, C.S. A Vector Space Model for Automatic Indexing. *Commun. ACM* **1975**, *18*, 613–620.
34. Manning, C.D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2008.
35. Liu, T.-M. Using Snake Roadkill Patterns to Indicate Effects of Climate Change on Snakes in Taiwan. *Sustainability* **2025**, *17*, 1580. <https://doi.org/10.3390/su17041580>
36. Jung, W.-C.; Kim, J.; Park, N. Web-Browsing Application Using Web Scraping Technology in Korean Network Separation Application. *Symmetry* **2021**, *13*, 1550. <https://doi.org/10.3390/sym13081550>
37. Mabrouk, A.; Redondo, R.P.D.; Kayed, M. SEOpinion: Summarization and Exploration of Opinion from E-Commerce Websites. *Sensors* **2021**, *21*, 636. <https://doi.org/10.3390/s21020636>
38. Chang, C.-W. Developing a Multicriteria Decision-Making Model Based on a Three-Layer Virtual Internet of Things Algorithm Model to Rank Players' Value. *Mathematics* **2022**, *10*, 2369. <https://doi.org/10.3390/math10142369>
39. Naganawa, H.; Hirata, E. Enhancing Policy Generation with GraphRAG and YouTube Data: A Logistics Case Study. *Electronics* **2025**, *14*, 1241. <https://doi.org/10.3390/electronics14071241>
40. Sergiacomi, C.; Vuletić, D.; Paletto, A.; Barbierato, E.; Fagarazzi, C. Exploring National Park Visitors' Judgements from Social Media: The Case Study of Plitvice Lakes National Park. *Forests* **2022**, *13*, 717. <https://doi.org/10.3390/f13050717>
41. Arreeras, S.; Phonsitthangkun, S.; Arreeras, T.; Arimura, M. Spatial Analysis on the Service Coverage of Emergency Facilities for Fire Disaster Risk in an Urban Area Using a Web Scraping Method: A Case Study of Chiang Rai City, Thailand. *Urban Sci.* **2024**, *8*, 140. <https://doi.org/10.3390/urbansci8030140>
42. Santos Duarte, A.L.; Souza, E.R.d.; Silva, M.P.d.O.; Monte, M.B.d.S.; Correia, N.O.d.A.; de Carvalho, V.D.H.; Taques, F.H. Data Protection in Brazil: Applying Text Mining in Court Documents. *Eng. Proc.* **2025**, *87*, 57. <https://doi.org/10.3390/engproc2025087057>
43. Pérez, V.; Aybar, C. Challenges in Geocoding: An Analysis of R Packages and Web Scraping Approaches. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 170. <https://doi.org/10.3390/ijgi13060170>

**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.