Analysis of Google News coverage: A comparative study of Brazil, Colombia, Mexico, Portugal, and Spain*

Dr. Douglas Farias Cordeiro
https://orcid.org/0000-0002-5187-0036
Universidade Federal de Goiás, Brazil
cordeiro@ufg.br

Dr. Carlos Lopezosa
http://orcid.org/0000-0001-8619-2194
Universitat de Barcelona, Spain
lopezosa@ub.edu

Dr. Javier Guallar
http://orcid.org/0000-0002-8601-3990
Universitat de Barcelona, Spain
jguallar@ub.edu

Dr. Mari Vállez
http://orcid.org/0000-0002-3284-2590
Universitat de Barcelona, Spain
marivallez@ub.edu

ABSTRACT. This study aims to examine the news coverage provided by Google News across five Ibero-American countries, including three from Latin America and two from Europe: Brazil, Colombia, Mexico, Portugal, and Spain. The main focus is to highlight the differences and similarities in news presentation within diverse contexts, evaluating the presence and distribution of news using quantitative indicators, and analyzing the predominant content in each country's news, based on a dataset collected between January 2 and January 31, 2024. This includes examining news sources, geographic coverage, prominent figures, and the prevalence of sensational elements through the identification of clickbaits. Our research employs statistical analyses and algorithmic solutions in

natural language processing and artificial intelligence to generate results. The analyses revealed consistency in the daily delivery of content by Google News, with specific variations in the update rate across the studied countries. A diversity of news sources was observed, with a greater tendency toward local news and frequent mentions of politicians, celebrities, and businesspeople. In addition, the analyses uncovered a significant presence of clickbait, with variations across the countries and topics.

KEYWORDS: Google News / digital media / news / news aggregators / data analysis

ANÁLISIS DE LA COBERTURA DE GOOGLE NEWS: UN ESTUDIO COMPARATIVO DE BRASIL, COLOMBIA, MÉXICO, PORTUGAL Y ESPAÑA

RESUMEN. Este estudio tiene como objetivo examinar la cobertura de noticias proporcionada por Google News en cinco países iberoamericanos, incluidos tres de América Latina y dos de Europa: Brasil, Colombia, México, Portugal y España. El enfoque principal es resaltar las diferencias y similitudes en la presentación de noticias en diversos contextos, evaluando la presencia y distribución de noticias mediante indicadores cuantitativos, y analizando el contenido predominante en las noticias de cada país, considerando un conjunto de datos extraído entre el 2 de enero y el 31 de enero de 2024. Esto incluye examinar las fuentes de noticias, la cobertura geográfica, las figuras prominentes y la prevalencia de elementos sensacionalistas a través de la identificación de prácticas de *clickbait*. Nuestra investigación emplea análisis estadísticos y soluciones algorítmicas en procesamiento del lenguaje natural e inteligencia artificial para generar resultados. Los análisis revelaron consistencia en la entrega diaria de contenido por parte de Google News, con variaciones específicas en la tasa de actualización entre los países estudiados. Se observó una diversidad de fuentes de noticias, con una mayor tendencia hacia las noticias locales y menciones frecuentes de políticos, celebridades y empresarios. Además, los análisis revelaron una presencia considerable de *clickbait*, con variaciones entre los países y los temas abordados.

PALABRAS CLAVE: Google News / medios digitales / noticias / agregadores de noticias / análisis de datos

ANÁLISE DA COBERTURA DO GOOGLE NEWS: UM ESTUDO COMPARATIVO DE BRASIL, COLÔMBIA, MÉXICO, PORTUGAL E ESPANHA

RESUMO. Este estudo tem como objetivo examinar a cobertura de notícias fornecida pelo Google News em cinco países ibero-americanos, incluindo três da América Latina e dois da Europa: Brasil, Colômbia, México, Portugal e Espanha. O principal foco é destacar as diferenças e semelhanças na apresentação das notícias em diversos contextos,

avaliando a presença e distribuição de notícias por meio de indicadores quantitativos, e analisando o conteúdo predominante nas notícias de cada país, considerando um conjunto de dados extraído entre 2 de janeiro e 31 de janeiro de 2024. Isso inclui examinar as fontes de notícias, a cobertura geográfica, as figuras proeminentes e a prevalência de elementos sensacionalistas através da identificação de práticas de *clickbait*. Nossa pesquisa emprega análises estatísticas e soluções algorítmicas em processamento de linguagem natural e inteligência artificial para gerar resultados. As análises revelaram consistência na entrega diária de conteúdo pelo Google News, com variações específicas na taxa de atualização entre os países estudados. Observou-se uma diversidade de fontes de notícias, com uma maior tendência para as notícias locais e menções frequentes a políticos, celebridades e empresários. Além disso, as análises revelaram uma presença considerável de *clickbait*, com variações entre os países e os temas abordados.

PALAVRAS-CHAVE: Google News / mídia digital / notícias / agregadores de notícias / análise de dados

INTRODUCTION

Google News is a free service provided by Google that aggregates daily headlines from hundreds of news sources, presenting them in a format resembling an online news outlet and offering different and independent local editions in a large number of countries worldwide (Google, 2024). Technically, Google News functions as a news aggregator, being one of the most prominent systems on the web that both professionals and the public use to access news content. Other similar systems include media outlets themselves, search engines, tracking platforms, news databases, and news curation through social media (Guallar et al., 2013; Vermeer et al., 2020).

The most popular news aggregators today, in addition to Google News, include Google Discover, Flipboard, Upday, News Republic, and Feedly (Negredo, 2023; Park, 2022). It is estimated that Google News generates approximately six billion clicks per month to publishers worldwide (Patel, 2019). Furthermore, it contributes to the development of qualified traffic that media companies can aim to retain. In fact, various studies suggest that 60 % of users trust Google News as a reliable source of information (Wilson, 2021).

Given the strategic importance of this service for online media, in addition to its widespread use by broad audiences, this study aims to research and compare the news coverage provided by Google News across five Ibero-American countries, including three from Latin America and two from Europe: Brazil, Colombia, Mexico, Portugal, and Spain. It highlights the similarities and differences in the presentation of news in different contexts. From this main objective, the following specific objectives emerge:

SO1: To analyze the presence and distribution of news across five Google News editions (Brazil, Colombia, Mexico, Portugal, and Spain) in terms of daily volume and update rate.

SO2: To analyze the predominant news in each country in terms of news sources, geographic coverage, topics, and individuals.

SO3: To analyze the presence of sensationalism in each country's news through the identification of clickbaits.

This comparative analysis seeks to offer valuable insights into how the Google News platform addresses and reflects news events across a group of countries.

LITERATURE REVIEW

A significant corpus of academic research has focused on Google News from various perspectives. These approaches cover topics such as news dissemination and consumption, its impact on public opinion, news personalization, coverage of health news, information retrieval and natural language processing, and finally, its relationship with the media (Lopezosa et al., 2024).

In research concerning the dissemination of news through Google News, its consumption and its impact on readers' decisions, prominent studies have centered on analyzing the geographic coverage of this Google service, identifying the affiliated media outlets and the types of news featured on the front page. These analyses encompass specific case studies in countries such as the United States and India (Watanabe, 2013), Germany (Schroeder & Kralemann, 2005), Brazil, Colombia, and Mexico (Cobos, 2020, 2021), as well as Spain and Brazil (Cordeiro, 2024), among others. Additionally, these studies have explored news reception through responses to specific searches on Google News, including queries related to local news (Fischer et al., 2020) or specific terms such as "information seeking" (Wilson & Maceviciute, 2013).

In this context, several studies have explored the impact of Google News on public opinion and, ultimately, on readers' decisions. Noteworthy are works that examined how information acquired through Google News influences financial decisions (Du & Song, 2022), and research that analyzed the international news section of Google News to evaluate the agenda set by this aggregator and its effect on readers (Young & Atkin, 2022).

Regarding research on Google News and its personalization, studies have examined various aspects of this topic, such as evaluating the degree of news personalization generated by news search algorithms (Evans et al., 2022), exploring the effects of implicit and explicit personalization on content and source diversity (Haim et al., 2018), analyzing the level of web search personalization to assess the risk of trapping users in the so-called "filter bubbles" (Cozza et al., 2016), and examining search results in Google News that adapt to each user's browsing history (Le et al., 2019).

These studies have generally sought to address crucial issues surrounding news search personalization and source diversity (Evans et al., 2022), examine how Google News' explicit personalization affects the diversity of presented news (Haim et al., 2018), measure the level of personalization in the results returned by Google News to different types of users within a news dataset (Cozza et al., 2016), and finally, analyze search results to quantify the extent and conditioning of personalization of these results within this Google service (Le et al., 2019).

Another predominant area of study on Google News pertains to the coverage of medical news and its aggregation within the Google News service. Notable research in this domain has focused on specific medical topics, including colorectal cancer (Basch et al., 2022), mammography (Young-Lin & Rosenkrantz, 2017), and breastfeeding (Seror et al., 2010).

Although each of these studies has its own central point, they all share a common denominator: the study of Google News and the news it presents in relation to their respective topics.

Research on the relationship between Google News and the media has focused on two areas. First, it seeks to understand the experiences and perceptions of chief editors, directors, and media owners indexed in Google News, exploring their views and experiences with the aggregator (Cobos, 2018). Second, it analyzes the application of regulations such as in the case of Spain's Intellectual Property Law (Guallar, 2015). This includes the suspension of the Google News service in Spain and the impact on Spanish online media after its closure in 2014 (Calzada & Gil, 2020).

Finally, a substantial body of prominent studies on Google News centers on information retrieval, covering controlled vocabularies, algorithms, and technical aspects that examine the platform in relation to the development of products and protocols for information search and aggregation. These studies explore various areas, such as learning for information filtering and data mining (Montejo-Ráez et al., 2009, 2010), algorithms and natural processing within Google News (Wubben et al., 2011), natural language processing specifically focused on Chinese (Hong et al., 2006, 2009) and Arabic (Alzahrani, 2013), news flow (Das et al., 2007), and text mining and social networks (Joshi & Gatica-Perez, 2006).

After illustrating the main research areas concerning Google News from 2005 to the present, this work is deemed capable of complementing and expanding some of the lines developed so far by conducting a comparative analysis of the news coverage provided by Google News across different countries.

MATERIALS AND METHODS

This study uses a descriptive and quantitative approach to examine the news aggregator Google News across five Ibero-American countries, including three from Latin America and two from Europe: Brazil, Colombia, Mexico, Portugal, and Spain. Two distinct web scraping procedures (Mitchell, 2024) were employed. First, the entire homepage of the Google News editions for each of the five countries was captured. Second, a separate web scraping process specifically targeted the Top Stories section, which is accessible via a dedicated link at the top of the homepage. The homepage and the Top Stories section of the Google News editions of the aforementioned countries were analyzed over a thirty-day period, from January 2 to January 31, 2024. The web scraping process was conducted once daily. The data were integrated into a structured database featuring the following attributes: news headline, media outlet, and publication date. The data obtained were processed considering their format and values at the source, i.e., assuming the reliability of Google News' procedures for aggregating, organizing, and making information available.

The data underwent enrichment processes aiming to generate the following derived indicators: geographic coverage, names of individuals, and clickbait tags. Statistical calculation procedures and artificial intelligence techniques were employed, particularly in the field of natural language processing (NLP) and pretrained language models (PLMs).

A quantitative analysis was performed to extract meaningful insights through a tabular format and data visualization approach. Five dimensions were considered for this analysis: daily distribution of aggregated news during the analyzed period, most recurrent news sources, geographic coverage of the news, occurrence of individual names, and identification of clickbait.

The first analyzed dimension focused on evaluating the distribution of news. Measures of central tendency—including total accumulated news items, mean, and standard deviation—were employed in this evaluation. A relevant aspect of this analysis pertained to verifying the update rate, that is, the timeliness of the delivered news. However, the existing literature lacks a specific definition of metrics for this purpose. Therefore, we propose a metric based on a statistically calculated penalty, which considers the time difference, in days, between the publication of a news item in its source and its indexing within Google News. This metric is defined as follows:

$$u_i = 1.0 - \begin{cases} 0.0, & \text{if } (d_g - d_n = 0) \\ 0.1, & \text{if } (d_g - d_n = 1) \\ 0.2, & \text{if } (d_g - d_n = 2) \\ 0.3, & \text{if } (d_g - d_n = 3) \\ 0.4, & \text{if } (4 \leq d_g - d_n \leq 7) \\ 0.5, & \text{if } (8 \leq d_g - d_n \leq 14) \\ 0.75, & \text{if } (15 \leq d_g - d_n \leq 21) \\ 0.95, & \text{if } (22 \leq d_g - d_n \leq 30) \\ 1.0, & \text{if } (d_g - d_n > 30) \end{cases}$$

where u_i represents the update rate, d_g is the date of indexing in Google News, and d_n is the date of publication of the news in the source.

The analysis of news sources aggregated by Google News was crucial for providing insights into the diversity and predominance of the included media outlets. This, in turn, revealed trends related to content patterns or editorial biases characteristic of each source and their impact on the visibility and dissemination of specific types of news. The news sources were directly identified during data extraction, with their frequency calculated for each Google News edition. During the source identification process, an automated data validation check was conducted by cross-referencing the source names with their corresponding link, in order to assess the correctness of Google News' tagging. This verification procedure was carried out using an algorithm based on regular expressions and Levenshtein distance.

To identify the geographic coverage, an ensemble learning strategy was employed, based on a pre-trained multilingual BERT (bidirectional encoder representations

from transformers) model (Devlin et al., 2019) and named entity recognition (NER) (Li et al., 2022). The algorithm identified occurrences of cities and countries in the news headlines, assigning the article's origin accordingly. If the country of origin was not detected, the news was labeled with the country corresponding to the Google News being analyzed. Through validation against a pre-labeled sample, the solution achieved an accuracy of 96.50 %. This procedure did not aim to classify the geographic location of the news outlet itself but rather to identify potential geographic references within the news content based on an analysis of its headline.

The identification of people's names in news headlines allowed for tracking the most frequently mentioned personalities in Google News' coverage. A pre-trained BERT model was used to recognize names. In a post-processing stage, variations referring to the same person (e.g., "Trump" and "Donald Trump") were consolidated using a country-specific predefined association dictionary, which also includes acronyms (e.g., "AMLO" for Andrés Manuel López Obrador). This association dictionary was created through empirical observations of the news articles comprising the database. The accuracy measured during the validation procedure reached 95.40 %.

Identifying clickbaits contributed to assess the quality of information provided by the news. It is common for headlines designed to attract clicks to not necessarily offer contextual content or accurate information. Furthermore, a high prevalence of sensationalist headlines can indicate low-quality sources, affecting the audience's experience and trust in aggregation platforms (Fu et al., 2017). This study employed a multilingual model based on the PLM xlm-roberta-large (Conneau et al., 2020) for clickbait detection, with an accuracy of 97.59 % reported in its original tests (Christodoulou, 2024). Subsequently, a boolean label was assigned to each of the news articles extracted from Google News for further analysis.

Additionally, to better understand the occurrence of clickbait, a similarity graph visualization was employed. Similarity graphs are graphical representations that depict relationships between words or categories based on their co-occurrences within a text corpus, helping to detect words that frequently appear together and reveal semantic or thematic associations. In content analysis, similarity graphs are useful for understanding lexical networks and identifying clusters of terms that share conceptual proximity. Using statistical methods such as Vergès' similarity analysis, the connections between terms are displayed in a network format, where nodes represent words and edges indicate the strength of their co-occurrence (Veremyev et al., 2019).

The extracted data were processed in their original format, as captured directly from Google News, without prior curation. This approach relied on the platform's automated aggregation and indexing processes, assuming the accuracy of Google News' internal mechanisms. However, the focus on five Ibero-American countries limits the generalizability of the results, leaving broader trends and regional variations in news aggregation

beyond this context unexplored. Moreover, although pretrained language models—such as those used for named entity recognition and clickbait detection—demonstrate high accuracy, they may be prone to errors arising from linguistic variations, particularly in multilingual environments. Despite these limitations, the study offers a rigorous quantitative analysis that seeks to identify patterns and trends in news aggregation.

RESULTS

Through the implementation of data extraction procedures, two initial datasets were generated: the first comprises the Google News homepages, totaling 5,099 news headlines, while the second pertains to the Top Stories section, with 38,333 news headlines. The two extracted datasets were analyzed separately in order to enhance the inherent characteristics of each examined Google News edition. All data underwent attribute treatment and enrichment procedures, aligned with the analyses across the five dimensions considered.

Daily Distribution

Table 1 shows the indicators related to the data extracted from the homepages, segmented by country. When examining the standard deviation, a comparable volume of aggregated news is observed across all countries, with the maximum number of aggregated news reaching 37 for each country. Concerning the update rate, Spain exhibits the highest value, closely approaching the maximum possible value. This indicates a significant daily delivery of news generated at the source on the same day it is aggregated by Google News. Conversely, Brazil, Colombia, Mexico, and Portugal exhibit an update rate of approximately 96 %, displaying similar variation in terms of standard deviation.

Table 1

General Statistics of the Homepages

	Brazil	Colombia	Mexico	Portugal	Spain
Average Daily Volume (Standard Deviation - SD)	35.93 (1.96)	32.77 (2.60)	33.67 (2.32)	33.83 (2.27)	33.73 (1.76)
Update Index (SD)	96.27 % (1.83 %)	96.67 % (1.72 %)	96.26 % (2.01 %)	96.22 % (2.10 %)	98.78 % (1.08 %)
Minimum Daily Value	29	28	28	27	30
Maximum Daily Value	37	37	37	37	37

Source: Own elaboration

Table 2 presents the trend indicators for the dataset of news aggregated on the headlines pages. Unlike the data from the homepages, a notable variation was observed

in Google News Portugal in terms of average volume and standard deviation. While Brazil, Colombia, Mexico, and Spain showed minimum aggregation values ranging from 236 to 252, Portugal had a minimum daily news delivery occurrence equal to 70, which is reflected in a high standard deviation.

Regarding the update rate of the Top Stories section, Brazil and Spain exhibited similar values, around 96 %, while Colombia, Mexico, and Portugal ranged between 94 % and 95 %. Although this percentage difference may seem small, a lower value indicates a higher occurrence of older news or a tendency toward a lack of "refreshment" in the aggregation of new content, meaning that the same news remains aggregated for more than a day.

Table 2

General Statistics of the Top Stories Section

	Brazil	Colombia	Mexico	Portugal	Spain
Average Daily Volume (Standard Deviation - SD)	262.50 (13.00)	258.97 (12.86)	267.67 (9.84)	229.23 (41.23)	254.40 (12.30)
Update Index (SD)	96.13 % (1.85 %)	94.67 % (2.26 %)	95.01 % (1.84 %)	94.62 % (2.61 %)	96.62 % (1.50 %)
Minimum Daily Value	244	236	252	70	238
Maximum Daily Value	308	298	309	258	293

News Sources

Table 3 presents the top ten sources aggregated on the homepages, detailing both the total numerical values and their respective percentage in relation to the dataset. A total of 162 sources were identified for the Brazil edition, 93 for Colombia, 124 for Mexico, 92 for Portugal, and 153 for Spain.

Table 3

Top Ten Sources on the Homepages

Brazil		Colombia		Mexico		Portugal		Spain	
G1	157 (14.56 %)	El Tiempo	139 (14.14 %)	El Universal	163 (16.14 %)	Público	117 (11.52 %)	El PAÍS	95 (9.39 %)
UOL Confere	63 (5.84 %)	Revista Semana	122 (12.41 %)	El Financiero	68 (6.73 %)	RTP Notícias	90 (8.86 %)	elDiario.es	68 (6.72 %)
Poder360	42 (3.90 %)	El Colombiano	109 (11.09 %)	Milenio	58 (5.74 %)	Expresso	66 (6.50 %)	20minutos.es	62 (6.13 %)
Terra	34 (3.15 %)	El Espectador	66 (6.71 %)	CNN en Español	49 (4.85 %)	Dioguinho	55 (5.41 %)	El Mundo	52 (5.14 %)
Aos Fatos	33 (3.06 %)	Pulzo.com	53 (5.39 %)	La Jornada	44 (4.36 %)	Notícias ao Minuto	49 (4.82 %)	ABC.es	48 (4.74 %)
AFP.com	32 (2.97 %)	CNN en Español	49 (4.98 %)	Periódico Excélsior	37 (3.66 %)	Observador	48 (4.72 %)	RTVE	44 (4.35 %)
Metrópoles	31 (2.88 %)	W Radio	35 (3.56 %)	El Heraldo de México	33 (3.27 %)	Diário de Notícias	46 (4.53 %)	Cadena SER	38 (3.75 %)
UOL	31 (2.88 %)	Caracol Radio	29 (2.95 %)	Uno TV Noticias	32 (3.17 %)	O MINHO	41 (4.04 %)	La Razón	36 (3.56 %)
Globo.com	29 (2.69 %)	El País Cali	25 (2.54 %)	Aristegui Noticias	26 (2.57 %)	Jornal de Notícias	38 (3.74 %)	El Periódico	29 (2.87 %)
CartaCapital	25 (2.32 %)	Portafolio	23 (2.34 %)	DW (Español)	25 (2.48 %)	Polígrafo	31 (3.05 %)	Lecturas	27 (2.67 %)

In Google News Brazil indicators, a high concentration is observed in G1, which holds a 14.56 % share. G1 is a Brazilian news portal owned by Grupo Globo, considered one of the leading traditional media outlets in the country. A noteworthy aspect of Brazil's data is that three of the top ten sources are related to fact-checking (UOL Confere, Aos Fatos, and AFP.com). This concentration pattern observed in Google News Brazil also appears similarly in the editions for Colombia, Mexico, and Portugal. In Google News Colombia, three legacy media dominate: *El Tiempo*, *Revista Semana*, and *El Colombiano*. In Google News Mexico, *El Universal* stands out with the highest percentage share (16.14 %), even compared to all other editions. It is noteworthy that the Colombian and Mexican editions include international media among the top ten sources, such as the US-based CNN en Español and the Latin American edition of Germany's DW. In Google News Portugal, Dioguinho—a media outlet specializing in reality TV shows and gossip—stands out. In Google News Spain, unlike the other editions, the percentage differences among the leading media are more evenly distributed, with all shares below 10 %.

For the Top Stories section, the following total number of sources were identified: Brazil 598, Colombia 210, Mexico 320, Portugal 293, and Spain 460. Table 4 presents the top ten sources according to the recurrence of news aggregation. The Brazilian, Colombian, and Mexican editions of Google News showed a concentration in a specific source, each surpassing a 10 % share, with the same media occupying the top spot on the homepage indicators. Meanwhile, the Portuguese and Spanish editions showed a lower percentage variation among the top sources, with differences in the main sources.

Table 4
Top Ten Sources in the Top Stories Section

Brazil		Colombia		Mexico		Portugal		Spain	
G1	1207 (15.33 %)	El Tiempo	1052 (13.54 %)	El Universal	889 (11.07 %)	A Bola	540 (7.85 %)	El Mundo	513 (6.59 %)
UOL Confere	483 (6.13 %)	El Colombiano	706 (9.09 %)	Milenio	498 (6.20 %)	Público	537 (7.81 %)	La Razón	439 (5.64 %)
UOL	391 (4.97 %)	Revista Semana	697 (8.97 %)	El Financiero	413 (5.14 %)	Notícias ao Minuto	437 (6.35 %)	20minutos.es	431 (5.54 %)
Poder360	335 (4.25 %)	El Espectador	530 (6.82 %)	La Jornada	384 (4.78 %)	RTP Notícias	371 (5.39 %)	EL PAÍS	350 (4.50 %)
Metrópoles	289 (3.67 %)	Caracol Radio	417 (5.37 %)	Uno TV Noticias	329 (4.10 %)	Expresso	349 (5.07 %)	ABC.es	313 (4.02 %)
O Antagonista	280 (3.56 %)	W Radio	342 (4.40 %)	Diario Deportivo Récord	285 (3.55 %)	Record	319 (4.64 %)	elDiario.es	264 (3.39 %)
Terra	248 (3.15 %)	Pulzo.com	266 (3.42 %)	CNN en Español	260 (3.24 %)	O Jogo	317 (4.61 %)	Europa Press	216 (2.78 %)
R7.com	177 (2.25 %)	FutbolRed	248 (3.19 %)	Periódico Excélsior	244 (3.04 %)	Jornal de Notícias	309 (4.49 %)	La Vanguardia	187 (2.40 %)
Estado de Minas	143 (1.82 %)	AS Colombia	242 (3.11 %)	Aristegui Noticias	191 (2.38 %)	Observador	290 (4.22 %)	Heraldo.es	184 (2.36 %)
O Tempo	143 (1.82 %)	Blu Radio	235 (3.02 %)	Mediotiempo	190 (2.37 %)	SIC Notícias	270 (3.93 %)	Cadena SER	178 (2.29 %)

Source: Own elaboration

As previously observed on the homepages, Google News Brazil's headlines pages featured among the top sources a fact-checking specialized media outlet, UOL Confere, which held a share of 6.13 %. The Colombian edition, in contrast to the homepages, did not feature international sources among the top ten. On the other hand, Google News Mexico maintained the presence of CNN en Español. In Google News Portugal, the top spot was occupied by the sports news outlet A Bola, while another sports-oriented outlet, O Jogo, was also observed. For Spain, there was a slight comparative variation from the home-page indicators, with mainly Spanish-origin sources.

Geographic Distribution of News

Table 5 presents the geographic distribution of news by countries for the homepages, based on the identification performed by the geographic mentions algorithm. There is a clear increase in news associated with each respective country, with Brazil, Colombia, and Spain exhibiting values above 80 %, while Mexico and Portugal show values of 78.91 % and 74.41 %, respectively. It is interesting to note that the Portuguese edition of Google News features a significant amount of news from another country, Brazil, with 11.71 %. This finding highlights the strong cultural, linguistic, and historical ties between Portugal and Brazil, which may influence the prominence of Brazilian news in the Portuguese edition of Google News. The shared language facilitates content exchange between the two nations (Müller et al., 2023), and Brazil's large media presence and influence in the lusophone world likely contribute to this high percentage, especially considering the significant number of Brazilian immigrants in Portugal. Additionally, it is worth noting that Brazil's larger territorial size compared to Portugal and, consequently, its greater number of media outlets, may result in a higher volume of aggregated Brazilian news sources in the Portuguese language. In contrast, the second largest participation in the other editions remains below 5 %.

Table 5
Top Five Countries on the Homepages

	Brazil		Colombia		Mexico		Portugal		Spain	
	Brazil	899 (83.40 %)	Colombia	805 (81.59 %)	Mexico	797 (78.91 %)	Portugal	756 (74.41 %)	Spain	835 (82.51 %)
	Israel	28 (2.60 %)	Ecuador	34 (3.46 %)	United States	39 (3.86 %)	Brazil	119 (11.71 %)	Israel	48 (4.74 %)
	United States	27 (2.50 %)	United States	18 (1.83 %)	Japan	37 (3.66 %)	Yemen	21 (2.07 %)	United States	15 (1.48 %)
	Japan	18 (1.67 %)	Switzerland	14 (1.42 %)	Ecuador	24 (2.38 %)	United States	16 (1.57 %)	France	11 (1.09 %)
	Yemen	18 (1.67 %)	Israel	14 (1.42 %)	Israel	17 (1.68 %)	Israel	13 (1.28 %)	Iran	11 (1.09 %)

In all editions, the United States is mentioned among the main countries in terms of percentage participation on the homepages. Another important aspect is the consistent appearance of Israel among the top five countries mentioned in the news across all editions, which is related to events concerning the Israeli–Palestinian conflict. Additionally, the presence of Japan stands out in the Brazilian and Mexican editions, associated with events stemming from an earthquake that occurred in early January 2024.

For the Top Stories section (Table 6), the percentage participation remains above 80 % for Brazil, Colombia, and Spain, as seen on their homepages, with Mexico also increasing to a share of 81.28 %. On the other hand, Google News Portugal showed a decrease from 74.41 % participation on the homepages to 71.54 % in the headlines pages, although news mentioning Brazil increased to 14.48 %. The presence of Israel continues to be significant across all analyzed editions. It is also noteworthy that Ukraine, in reference to the Russian–Ukrainian conflict, appears in the news of the Brazilian, Mexican, Portuguese, and Spanish editions.

Table 6

Top Five Countries in the Top Stories Section

Brazil		Colombia		Mexico		Portugal		Spain	
Brazil	6920 (87.87 %)	Colombia	6676 (85.93 %)	Mexico	6527 (81.28 %)	Portugal	4920 (71.54 %)	Spain	6744 (86.66 %)
United States	139 (1.77 %)	Ecuador	125 (1.61 %)	United States	294 (3.66 %)	Brazil	996 (14.48 %)	Israel	183 (2.35 %)
Israel	120 (1.52 %)	Mexico	110 (1.42 %)	Israel	186 (2.32 %)	Ukraine	88 (1.28 %)	Ukraine	95 (1.22 %)
Argentina	69 (0.88 %)	Spain	102 (1.31 %)	Spain	160 (1.99 %)	Spain	83 (1.21 %)	United States	88 (1.13 %)
Ukraine	62 (0.79 %)	Israel	100 (1.29 %)	Ukraine	93 (1.22 %)	Israel	75 (1.09 %)	France	59 (0.76 %)

Source: Own elaboration

Mentioned Names

Regarding mentions of people's names on the homepages, the Brazilian, Colombian, Mexican, and Spanish editions prominently feature politicians, with the presidents of each country being the most frequently mentioned names, significantly outpacing others. However, in the Portuguese edition, the current president does not appear among the top five most mentioned names, although the names of politicians, businesspeople, and celebrities are still present. Notably, the most mentioned name in Portugal is Deputy Pedro Nuno Santos (Table 7).

Table 7

Top Five Mentioned Names on the Homepages

Brazil 490 (45.45 %)	Colombia 426 (43.34 %)	Mexico 416 (41.19 %)	Portugal 440 (43.31 %)	Spain 435 (42.98 %)
Lula 68 (13.88 %)	Gustavo (Petro) 77 (18.08 %)	(Donald) Trump 14 (3.37 %)	Pedro Nuno Santos 34 (7.73 %)	(Pedro) Sánchez 43 (9.89 %)
Bolsonaro 27 (5.51 %)	Piedad Córdoba 12 (2.82 %)	Chicharito Hernádez 12 (2.88 %)	Cristina Ferreira 27 (6.14 %)	Feijóo 20 (4.60 %)
(Donald) Trump 17 (3.47 %)	(Álvaro) Uribe 11 (2.58 %)	Ernestina Godoy 11 (2.64 %)	António Costa 24 (5.45 %)	(Carles) Puigde-mont 11 (2.53 %)
Zagallo 16 (3.27 %)	Carlos Fernando Galán 10 (2.35 %)	Álvarez Máynez 10 (2.40 %)	Miguel Albu-querque 23 (5.23 %)	Felipe (VI) 11 (2.53 %)
Boulos 12 (2.45 %)	(Javier) Milei 10 (2.35 %)	Carlos Bremer 10 (2.40 %)	André Ventura 19 (4.32 %)	Letizia 10 (2.30 %)
	Sofía Vergara 10 (2.35 %)	(Javier) Milei 10 (2.40 %)		
		(Joe) Biden 10 (2.40 %)		

Note. Country name (value and percentage of occurrence relative to total records). Mentioned name (value and percentage of occurrence in records with name mentions).

Source: Own elaboration

In the Top Stories section, there is a notable presence of politicians or individuals associated with politics. In Brazil, Mexico, Portugal, and Spain, only individuals from this category were identified among the top five mentions. In the Colombian edition, in addition to politicians, the names of footballers Luis Díaz and Arturo Vidal were also identified. It is also interesting to note that in Brazil, Mexico, Portugal, and Spain, Donald Trump's name appears among the top five. Furthermore, the results for the Mexican edition indicate that the name of Benito Juárez—known as the first Indigenous person to serve as president of the country—is among the most frequently mentioned. This is primarily due to headlines referencing public policy programs of the Mexican government that bear the statesman's name (Table 8).

Table 8

Top Five Mentioned Names in the Top Stories Section

Brazil 2664 (33.83 %)		Colombia 2859 (36.80 %)		Mexico 2944 (36.66 %)		Portugal 2988 (43.45 %)		Spain Spain: 2434 (31.28 %)	
Lula	410 (15.39 %)	Gustavo (Petro)	363 (12.70 %)	(Claudia) Shein-baum	106 (3.60 %)	Pedro Nuno Santos	148 (4.95 %)	(Pedro) Sánchez	254 (10.44 %)
Bolsonaro	180 (6.76 %)	(Carlos Fernando) Galán	76 (2.66 %)	(Benito) Juárez	86 (2.92 %)	Marcelo (Rebelo de Sousa)	79 (2.64 %)	Feijóo	77 (3.16 %)
(Donald) Trump	84 (3.15 %)	(Luis) Díaz	61 (2.13 %)	(Donald) Trump	74 (2.51 %)	(André) Ventura	75 (2.51 %)	(Isabel Díaz) Ayuso	59 (2.42 %)
(Javier) Milei	77 (2.89 %)	(Álvaro) Uribe (Vélez)	56 (1.96 %)	(Xóchitl) Gálvez	73 (2.48 %)	(Donald) Trump	64 (2.14 %)	Yolanda Díaz	48 (1.97 %)
(Ricardo) Lewan-dowski	62 (2.33 %)	(Arturo) Vidal	47 (1.64 %)	(Joe) Biden	62 (2.11 %)	(António) Costa	62 (2.07 %)	(Donald) Trump	45 (1.85 %)

Note. Country name (value and percentage of occurrence relative to total records). Mentioned name (value and percentage of occurrence in records with name mentions).

Clickbait Identification

For the homepages, the measurement from the clickbait identification algorithm revealed that the Portuguese edition of Google News had the lowest percentage at 22.15 %, which is relatively close to the indicators for Spain (27.27 %) and Brazil (28.29 %). In contrast, Colombia and Mexico showed higher values, with 37.33 % and 37.52 %, respectively. To understand the impact of the aggregation of news identified as clickbait, similarity graphs were generated for each edition of Google News, and sample excerpts of five news articles were selected according to the clickbait label characterization (Table 9).

Table 9

Clickbait Indicators on the Homepages

Similarity Graphs	Descriptive Sample Excerpts
Brazil	

BBB24: Mãe de Vanessa Lopes manda recado afiado e fala sobre saúde mental da filha: 'Está na melhor fase'; assista (Hugo Gloss)

Mapeamos quem queria dar o golpe em 8 de Janeiro, diz Lula (Poder360)

Safadeza pura! Marido aproveita descuido da mulher para brincar com "amante", mas é surpreendido; assista (BNews)

Homens brigam no meio da rua e se matam com a mesma arma; veja vídeo (A10+)

Lucas Pasin: Dedo na cara em camarote: Yasmin Brunet sabe fazer barraco e manter classe (Splash)

| **Colombia** | |

Así respondió Sofía Vergara a las críticas de Roy Barreras sobre las narconovelas (Revista Semana)

Esta es la millonaria cifra que ganaría Arturo Vidal en América de Cali (El Heraldo)

Esta sería la pena a la que se expone Nicolás Petro si es hallado culpable de lavado de activos y enriquecimiento ilícito (Revista Semana)

El triple 'chulo azul' llega a Whatsapp: esta será su función (RCN Radio)

Horóscopo: conoce las predicciones para tu signo en amor, salud y dinero HOY 13 enero (Terra Colombia)

(*contines*)

(*continued*)

Similarity Graphs	Descriptive Sample Excerpts

Mexico

Así se burló el Club América de las Chivas tras el regreso de Chicharito Hernández (infobae)

Esta es la función que compartirán los tres nuevos Samsung Galaxy S24 por primera vez (Andro4all)

¿Cómo vender mis monedas de 20 pesos en los bancos de México? Estos son los requisitos (El Heraldo Binario)

Por esta razón AMLO va contra la Ley de pensiones de 1997 de Zedillo (Polemón)

Así quedó el avión accidentado en un aeropuerto de Japón | Video (CNN en Español)

Portugal

Tudo a dar os parabéns! Famosos reagem ao recente amor de Cristina Ferreira, 16 anos mais novo (Flash)

Ex de João Monteiro ainda mantém fotos com o namorado de Cristina Ferreira (A Televisão)

Há outro "Big Brother" cá fora: Márcia Soares deixa de seguir Francisco Monteiro (Selfie)

Reação que vai dar que falar! Márcia Soares reage assim à entrevista de Francisco Monteiro no Goucha (Dioguinho)

As pistas que Pedro Nuno Santos deixou em 2018 sobre como quer transformar a economia (Público)

(*continues*)

(*continued*)

Similarity Graphs	Descriptive Sample Excerpts

Spain

Así se enteró Bertín Osborne del nacimiento del hijo de Gabriela Guillén (ABC.es)

Samsung la lía y mejora los Galaxy S24 con algo que los usuarios del iPhone desearán al instante (ADSLZone)

La impactante confesión de Olvido Hormigos sobre sus hijos y su vídeo sexual (La Razón)

Pilar Eyre destapa el posible movimiento que podría hacer la reina Letizia en unos días (Lecturas)

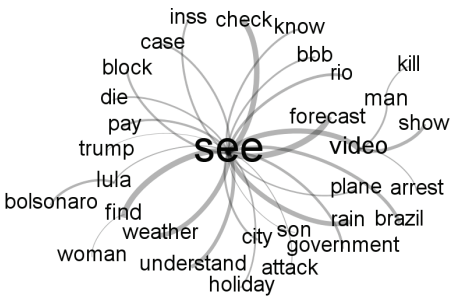Todo lo que se sabe de las medidas pactadas entre el PSOE y Junts para aprobar los decre-tos (elDiario.es)

Note. The words in the graphs have been translated into English, while the original language of the headlines has been retained.

On the homepages, it is observed that in Brazil, there is a thematic concentration of clickbait related to news about celebrities and television programs, highlighted by terms like "BBB" (Big Brother Brasil) and the names of entertainment personalities. Additionally, to a lesser extent, names of politicians such as Lula and Bolsonaro also appear. This trend is also notable in Portugal, where clickbait news essentially focus on entertainment and celebrity topics. In Colombia, although entertainment-related news remains prevalent, there is a greater thematic diversification, with more terms related to politics, business, and technology, similar to what is seen in Mexico. In contrast, Google News Spain shows a broader thematic diversification, evidenced by the dispersion of terms and a slight difference in the thickness of the graph edges, encompassing topics such as technology, politics, business, and entertainment.

In the Top Stories section, two distinct patterns emerge (Table 10). First, in percentage terms, Brazil, Portugal, and Spain present values below 20 %. Upon analyzing the main terms and selected samples, it is clear that the entertainment topic, which is prominent in the homepage indicators, does not have significant representativeness, with a tendency towards terms or names of individuals related to politics, sports, and society. Portugal stands out for its higher frequency of sports-related mentions. On the other hand, Colombia and Mexico present percentages around 30 %. In these regions, in addition to terms related to politics, sports, and society, there are mentions of economic issues and curiosities, indicating a broader diversification in terms of news types that make up the headlines.
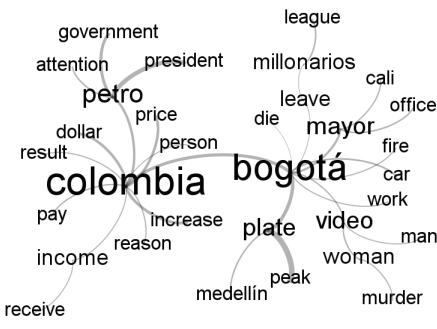
Table 10

Clickbait Indicators in the Top Stories Section

Similarity Graphs	Descriptive Sample Excerpts
Brazil	
	Veja o que fazer com os papéis da Vale (VALE3) após pressão de Lula e multa bilionária (Seu Dinheiro) Luxo e propina: quem é Maurício Demétrio, delegado condenado a nove anos de prisão no Rio (CartaCapital) Trump inelegível? Entenda decisões que podem tirar nome do magnata das eleições (O Tempo) Homem tenta furtar cabos se equilibrando em fios mas acaba caindo; veja vídeo (O Dia) Atleta passa mal durante corrida e morre com doença do xixi preto; entenda (O Tempo)

(continues)

(continued)

Similarity Graphs	Descriptive Sample Excerpts

Colombia

Precio dólar en Colombia dio giro inesperado: así quedó HOY viernes (RCN Radio)

Gobierno Petro ahora evalúa eliminar descuento del SOAT: estas son las razones (Valora Analitik)

VIDEO: escándalo de una mujer hizo que se retrasara un vuelo Bogotá-Medellín por más de dos horas. ¿Qué pasó? (Revista Semana)

Millonarios, a la espera: ¿qué pasa con la firma de su refuerzo Santiago Giordana? (El Tiempo)

Así adquirió el Distrito el primer coche eléctrico que ya rueda por Cartagena (El Universal)

Mexico

Así se burló el Club América de las Chivas tras el regreso de Chicharito Hernández (infobae)

Mujeres con Bienestar Edomex: esto sabemos del calendario de pagos para 2024 (infobae)

Chiefs vs. Dolphins fue el 4to partido con mayor frío en la historia: ¿A qué temperatura jugaron? (El Financiero)

Borracho destruye los coches de una periodista | Encima sus amigas la GOLPEAN (El Universal)

Ésta es la razón por la que Paola Suárez no quiere ver los videos de agresión que sufrió (Milenio)

(continues)

Similarity Graphs	Descriptive Sample Excerpts

Portugal

Sérgio Conceição não revela em quem votaria nas eleições do FC Porto (SAPO Desporto)

Marcos Leonardo agradeceu ao Benfica e até jogador do Real Madrid reagiu (A Bola)

Saiba quem é João Monteiro, o homem que faz o coração de Cristina Ferreira bater mais forte (Flash)

Al Nassr não esqueceu feito de Cristiano Ronaldo. Português reagiu assim (Notícias ao Minuto)

Tudo o que disse Rúben Amorim na antevisão ao jogo com o Tondela (A Bola)

Spain

Ordenan la retirada inmediata de este famoso queso en España y piden no consumirlo (EL ESPAÑOL)

¿Qué tiempo hará en Barcelona en lo que queda de enero? El Meteocat trae la peor de las noticias (MUNDO DEPORTIVO)

La actitud de Irene Urdangarin y Victoria Federica en el cumpleaños del rey Juan Carlos (20minutos.es)

Esto es lo que recomiendan los Bomberos de Madrid ante la inminente ola de frío (La Razón)

Así ha sido el discurso de Javier Milei en el Foro de Davos y la respuesta de Pedro Sánchez (20minutos.es)

Note. The words in the graphs have been translated into English, while the original language of the headlines has been retained.

DISCUSSION AND CONCLUSIONS

Google News has long been recognized as one of the most important web traffic-driving systems for media outlets (Young & Atkin, 2022). Alongside the news results from Google's general search engine (Giomelakis, 2023) and Google Discover's algorithmic news recommendation service (Lopezosa et al., 2024), it forms a perfect triad to attract more readers (Newman et al., 2023).

Research on web visibility—especially in relation to online media—is widespread, with a primary focus on analyzing—through reverse engineering—how news articles are ranked in featured results on platforms such as google.com, google.es, etc. (Lopezosa et al., 2019) and assessing their impact on both the journalistic enterprise (Pedrosa & de Morais, 2021) and readers (Evans et al., 2022). Our study aligns with this branch of research on web visibility but centers on Google News as the case study, adopting a descriptive, quantitative, and comparative approach across multiple countries.

Our study goes beyond previous research by examining Google News coverage across five countries, analyzing not only affiliated media outlets and news topics but also key individuals and the prevalence of sensationalism, identified through automated clickbait detection. Our findings affirm Google News' potential as a news aggregation service that enhances the online visibility of media outlets. This supports existing studies conducted in Germany (Schroeder & Kralemann, 2005), the United States, India (Watanabe, 2013), Brazil, Colombia, and Mexico (Cobos, 2018, 2021). We observed a substantial daily volume of news items on Google News, frequent updates, and a significant concentration of specific media outlets as primary sources of information.

These findings demonstrate, on the one hand, a competitive advantage in terms of web visibility for media outlets that frequently dominate the front page of Google News with a higher volume of news articles. On the other hand, they confirm some concerns about the degree of news personalization generated by search algorithms (Haim et al., 2018; Le et al., 2019; Evans et al., 2022). Specifically, they highlight the potential for this concentration of information sources on Google News front page to trap users in the so-called "filter bubbles" (Cozza et al., 2016).

This study introduces several innovations compared to other research on Google News, information retrieval, and algorithm use (Lopezosa et al., 2024). One key distinguishing feature of our study is its focus on the application of statistical calculation methods and artificial intelligence, particularly in the field of natural language processing and pretrained models. Our aim was to identify the thematic groups within the retrieved news, assess the degree of sensationalism, and highlight the main protagonists in the published information.

This methodological approach builds on a longstanding field of research that reached its peak between 2006 and 2013 and appears to be experiencing a resurgence,

partly due to advances in artificial intelligence (Cordeiro, 2024). Therefore, our work builds upon the dynamics of previous studies on Google News, learning for information filtering and data mining (Joshi & Gatica-Perez, 2006; Montejo-Ráez et al., 2009, 2010), natural language processing (Hong et al., 2006, 2009; Wubben et al., 2010; Alzahrani, 2013), and news flow (Das et al., 2007).

Our results suggest that up to a quarter of the analyzed news contains a degree of clickbait, confirming that attention-grabbing news holds significant weight within Google News. This phenomenon is also observed, for instance, in the Google Discover service (Lopezosa et al., 2024). However, specific studies are needed to further confirm this relationship.

We summarize the key findings of the study in relation to its objectives and research questions, evaluating the extent to which they were achieved. Additionally, we discuss the study's limitations and propose directions for future research. Our analysis of Google News coverage across different national contexts yields valuable insights into its dynamics and influence in aggregating, presenting, and disseminating news. By addressing our defined objectives and research questions, we draw conclusions across the five analyzed dimensions.

Regarding SO1 (news daily volume and update rate), first and foremost, the uniformity in the volume of news aggregated across the five analyzed editions of Google News stands out. The standard deviation reflects consistent delivery of daily informative content, both on the homepages and in the Top Stories section. Notably, the homepage for Spain exhibits a slightly higher update rate, indicating a prioritization of recent news. However, the other editions also show high update rates, with individual values around 96 % (Table 2). A slight variation was observed in the Top Stories section, where Brazil and Spain recorded higher update rates (96 %), while Colombia, Mexico, and Portugal fell just below 95 %.

Regarding SO2 (news sources, geographic coverage, individuals, and topics), each country displays a unique variety of news sources, reflecting the diversity of its local media landscape. In this dimension, certain sources stand out, such as G1 in Brazil, *El Universal* in Mexico, and A Bola in Portugal. Concerning geographic mentions and key individuals, a clear trend towards a predominance of local news is evident across all editions of Google News. In this context, the names of politicians, celebrities, and businesspeople are central to these mentions—with notable differences between country editions—reflecting public interest and editorial focus. Furthermore, the inclusion of international media outlets, such as CNN en Español and DW Español, demonstrates Google News' global reach. Additionally, some significant variations are observed in certain editions, such as Brazil's strong presence in Google News Portugal, which may reflect the historical and cultural ties between these two countries.

Furthermore, the consistent coverage of events concerning the United States and Israel in the news underscores their global significance and geopolitical impact. Local events, such as the earthquake in Japan, are also prominently featured in certain editions, illustrating the platform's responsiveness to significant national events from around the world. The diversity of topics and presentation styles across Google News editions reflects its adaptation to each country's cultural and informational context, potentially shaping public opinion and perceptions of global events.

Lastly, regarding SO3 (clickbait), while its presence varies across countries, certain recurring patterns are observed. These patterns reflect characteristics of the sources indexed by Google News, which are then reflected in the universe of aggregated news delivered to users. Topics related to entertainment and celebrities are particularly prominent in Brazil and Portugal, suggesting their strong appeal to users in these regions. In contrast, the editions for Colombia and Mexico exhibit a broader thematic diversity—ranging from politics and business to curiosities—indicating the platform's response to a wide array of interests and reflecting a greater adoption of sensationalist strategies in the face of the diverse audiences in these regions.

The comparative analyses have offered insights into how Google News interacts with news dynamics across Ibero-American countries. The similarities and differences among national editions underscore the platform's complexity as a significant player in aggregating, presenting, and providing access to information. These findings contribute to our understanding of digital media's influence in modern society, illustrating how news aggregation contributes to the global dissemination of information.

However, this study has certain limitations that should be acknowledged. The data collection was restricted to the homepage and the Top Stories section of Google News, focusing on a specific time period and a single daily collection. Such an approach may not adequately capture the dynamic content variations that occur in other sections or throughout the day. Additionally, the analyses rely on the accuracy of the artificial intelligence algorithms used for pattern recognition, such as clickbait detection and name identification, which may introduce slight noise into the findings. Finally, the Google News editions analyzed cover a limited set of Ibero-American countries, which restricts the generalizability of the results. Nevertheless, these limitations do not undermine the relevance of the study's conclusions.

Future research could involve expanding the analysis to identify specific topics in Google News' aggregated content using text analysis and natural language processing techniques. This approach would enhance our understanding of trends and thematic emphases across editions, offering insights into user interests. Additionally, broadening the sample in terms of time and including more editions from diverse countries would deepen our comprehension of variations and patterns in news coverage. Moreover, systematically comparing Google News with other aggregation tools like Google Discover,

Apple News, or Flipboard would contextualize its role and performance within the global digital media. These directions promise to provide comprehensive insights into the impact and operations of news aggregation platforms in today's digital era.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Conceptualization was conducted by D. F. C., C. L., and J. G.; data extraction by D. F. C.; analysis by D. F. C., C. L., J. G., and M. V.; research by D. F. C., C. L., and J. G.; methodology by D. F. C., C. L., and J. G.; original draft preparation by D. F. C., C. L., and J. G.; and review and editing by D. F. C., C. L., J. G., and M. V.

REFERENCES

Alzahrani, S. M. (2013). Building, profiling, analysing and publishing an Arabic news corpus based on Google News RSS feeds. In R. E. Banchs, F. Silvestri, T. Y. Liu, M. Zhang, S. Gao, & J. Lang –Eds.), *Lecture notes in computer science: Vol. 8281. Information retrieval technology* (pp. 488-499). Springer. https://doi.org/10.1007/978-3-642-45068-6_42

Basch, C. H., Hillyer, G. C., & Jacques, E. T. (2022). News coverage of colorectal cancer on Google News: Descriptive study. *JMIR Cancer*, *8*(2), Article e39180. https://doi.org/10.2196/39180

Calzada, J., & Gil, R. (2020). What do news aggregators do? Evidence from Google News in Spain and Germany. *Marketing Science*, *39*(1), 134-167. https://doi.org/10.1287/mksc.2019.1150

Christodoulou, C. (2024). *XLM-RoBERTa-Multilingual-Clickbait-Detection*. Hugging Face. https://huggingface.co/christinacdl/XLM_RoBERTa-Multilingual-Clickbait-Detection

Cobos, T. L. (2018). Perceptions and experiences about Google News from the editors of Latin-American news media indexed in the editions of Colombia and Mexico. *Estudios sobre el Mensaje Periodístico*, *24* (2), 1183-1198. https://hdl.handle.net/20.500.12585/9225

Cobos, T. L. (2020). Journalism industries in the internet era: The case of Colombian news media outlets in Google News Colombia. *Contratexto*, (33), 85-104. https://doi.org/10.26439/contratexto2020.n033.4785

Cobos, T. L. (2021). Origin and weight of news media outlets indexed on Google News: An exploration of the editions from Brazil, Colombia, and Mexico. *Brazilian Journalism Research, 17*(1), 28-63. https://doi.org/10.25200/BJR.v17n1.2021.1331

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. *Proceedings of the 58th annual meeting of the Association for Computational Linguistics* (pp. 8440-8451). https://doi.org/10.18653/v1/2020.acl-main.747

Cordeiro, D. F. (2024). Perspectivas en contraste: análisis comparativo cuantitativo España y Brasil de la cobertura del conflicto israelí-palestino en Google News: análise comparativa quantitativa Espanha e Brasil da cobertura do conflito israelo-palestino no Google News. *Documentación de las Ciencias de la información, 47*, 15-25. https://doi.org/10.5209/dcin.92187

Cozza, V., Hoang, V.T., Petrocchi, M., Spognardi, A. (2016). Experimental measures of news personalization in Google News. In S. Casteleyn, P. Dolog, & C. Pautasso. (Eds.), *Current trends in web engineering* (pp. 93–104). Springer. https://doi.org/10.1007/978-3-319-46963-8_8

Das, A. S., Datar, M., Garg, A., & Rajaram, S. (2007). Google news personalization: scalable online collaborative filtering. *WWW '07: Proceedings of the 16th international conference on World Wide Web, New York, NY, USA*, 271-280. https://doi.org/10.1145/1242572.1242610

Devlin, J., Chang, M., Lee, K, & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio. (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies, Volume 1* (pp. 4171-4186). Association for Computational Linguistics. https://doi.org/10.48550/arXiv.1810.04805

Du, K., & Song, J. (2022). The impact of geotargeting on household information acquisition: Evidence from a Google News redesign. *Research Policy, 51*(10), Article 104596. https://doi.org/10.1016/j.respol.2022.104596

Evans, R., Jackson, D., & Murphy, J. (2022). Google news and machine gatekeepers: Algorithmic personalisation and news diversity in online news search. *Digital Journalism, 11*(9), 1682-1700. https://doi.org/10.1080/21670811.2022.2055596

Fischer, S., Jaidka, K., & Lelkes, Y. (2020). Auditing local news presence on Google News. *Nature Human Behaviour, 4*(12), 1236-1244.

Fu, J., Liang, L., Zhou, X., & Zheng, J. (2017). A convolutional neural network for clickbait detection. In S. Li, Y. Dai, & Y. Cheng. (Eds.), *2017 4th International Conference on Information Science and Control Engineering (ICISCE)* (pp. 6-10). CPS. https://doi.org/10.1109/ICISCE.2017.11

Giomelakis, D. (2023). Semantic search engine optimization in the news media industry: Challenges and impact on media outlets and journalism practice in Greece. *Social Media + Society*, *9*(3). https://doi.org/10.1177/20563051231195545

Google. (2024). *Get started with Google News.* https://support.google.com/googlenews/answer/9005669?hl=en&co=GENIE.Platform%3DAndroid

Guallar, J. (2015). Prensa digital en 2013-2014. *Anuario ThinkEPI*, *9*, 153–160. https://doi.org/10.3145/thinkepi.2015.37

Guallar, J., Abadal, E., & Codina, L. (2013). Sistemas de acceso a la información de prensa digital: tipología y evolución. *Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información*, *27*(61), 29-52. https://doi.org/10.1016/S0187-358X(13)72553-X

Haim, M., Graefe, A., & Brosius, H. B. (2018). Burst of the filter bubble? Effects of personalization on the diversity of Google News. *Digital Journalism*, *6*(3), 330-343. https://doi.org/10.1080/21670811.2017.1338145

Hong, C., Chen, C., & Chiu, C. (2006). New word extraction utilizing Google News corpuses for supporting lexicon-based Chinese word segmentation systems. *The 2006 IEEE International Joint Conference on Neural Networks Proceedings*, *Vancouver, BC,* 3040-3046. https://doi.org/10.1109/IJCNN.2006.247263

Hong, C., Chen, C., & Chiu, C. (2009). Automatic extraction of new words based on Google News corpora for supporting lexicon-based Chinese word segmentation systems. *Expert Systems with Applications: An International Journal*, *36*(2), 3641-3651. https://doi.org/10.1016/j.eswa.2008.02.013

Joshi, D., & Gatica-Perez, D. (2006). Discovering groups of people in google news. *Proceedings of the 1st ACM International Workshop on Human-Centered Multimedia (HCM '06), New York, NY, USA,* 55-64. https://doi.org/10.1145/1178745.1178757

Le, H., Maragh, R., Ekdale, B., High, A., Havens, T., & Shafiq, Z. (2019). Measuring political personalization of Google News search. In L. Liu, & R. White (Eds.). *WWW '19: The Web Conference 2019* (pp. 2957-2963). Association for Computing Machinery. https://doi.org/10.1145/3308558.3313682

Li, J., Sun, A., Han, J., & Li, C. (2022). A survey on deep learning for named entity recognition: extended abstract. *IEEE Transactions on Knowledge and Data Engineering*, *34*(1), 50-70. https://doi.org/10.1109/TKDE.2020.2981314

Lopezosa, C., Codina L., & Rovira, C. (2019). *Visibilidad web de portales de televisión y radio en España*: ¿qué medios llevan a cabo un mejor posicionamiento en buscadores? Universitat Pompeu Fabra, Barcelona. https://repositori.upf.edu/handle/10230/36234

Lopezosa, C., Giomelakis, D., Pedrosa, L., & Codina, L. (2024). Google Discover: uses, applications and challenges in the digital journalism of Spain, Brazil and Greece. *Online Information Review*, *48*(1), 123-143. https://doi.org/10.1108/OIR-10-2022-0574

Lopezosa, C., Vállez, M., & Guallar, J. (2024). The vision of Google News from the academy: scoping review. *Doxa Comunicación, 38*, 317-332. https://doi.org/10.31921/doxacom.n38a1891

Mitchell, R. (2024). *Web scraping with Python: Data extraction from the modern web*. O'Reilly Media.

Montejo-Ráez, A., Perea-Ortega, J. M., Díaz-Galiano, M. C., & Ureña-López, L. A. (2009). SINAI at INFILE 2009: Experiments with Google News. *CEUR Workshop Proceedings: Vol. 1175*. https://ceur-ws.org/Vol-1175/CLEF2009wn-INFILE-MontejoRaezEt2009.pdf

Montejo-Ráez, A., Perea-Ortega, J. M., Díaz-Galiano, M. C., & Ureña-López, L. A. (2010). Experiments with Google News for filtering newswire articles. In C. Peters, G. M. Nunzio, M. Kurimo, T. Mandl, D. Mostefa, A. Peñas, & G. Roda (Eds.), *Lecture Notes in Computer Science: Vol. 6241. Multilingual Information Access Evaluation I. Text Retrieval Experiments* (pp. 381-384). Springer. https://doi.org/10.1007/978-3-642-15754-7_46

Müller, M. S., Cabecinhas, R., & Santos Silva, D. (2023). Cultural journalism in Brazil and Portugal: cross-country analysis. *Brazilian Journalism Research*, *19*(1), Article e1546. https://doi.org/10.25200/BJR.v19n1.2023.1546

Negredo, S. (2023). Uno de cada cuatro internautas españoles dice usar Google News, y uno de cada cinco, Discover. *Digital News Report España.* https://www.unav.edu/web/digital-news-report/entradas/-/blogs/uno-de-cada-cuatro-internautas-espanoles-dice-usar-google-news-y-uno-de-cada-cinco-discover

Newman, N., Fletcher, R., Eddy, K., Robertson, C. T., & Nielsen R. K. (2023). *Reuters Institute digital news report 2023*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2023-06/Digital_News_Report_2023.pdf

Park, C. S. (2022). Reading a snippet on a news aggregator vs. clicking through the full story: Roles of perceived news importance, news efficacy, and news-finds-me perception. *Journalism Studies*, *23*(11), 1350-1376. https://doi.org/10.1080/1461670X.2022.2086160

Patel, N. (2019). Cómo publicar tu sitio en Google News y generar más tráfico en tiempo real. *Neilpatel.com*. https://neilpatel.com/es/blog/como-publicar-tu-sitio-en-google-news-y-generar-mas-trafico-en-tiempo-real/

Pedrosa, L., & de Morais, O. J. (2021). Visibilidade web nos buscadores: Fatores algorítmicos de SEO on-page (FAOPs) como técnica e prática jornalística. *Estudios sobre el Mensaje Periodístico*, *27*(2), 579-591. https://doi.org/10.5209/esmp.71291

Schroeder, R., & Kralemann, M. (2005). Journalism ex Machina-Google News Germany and its news selection processes. *Journalism Studies*, *6*(2), 245-247. https://doi.org/10.1080/14616700500057486

Seror, J., Amar, A., Braz, L., & Rouzier, R. (2010). The Google News effect: Did the tainted milk scandal in China temporarily impact newborn feeding patterns in a maternity hospital? *Acta Obstetricia et Gynecologica Scandinavica*, *89*(6), 823-827. https://doi.org/10.3109/00016349.2010.484046

Veremyev, A., Semenov, A., Pasiliao, E. L., & Boginski, V. (2019). Graph-based exploration and clustering analysis of semantic spaces. *Applied Network Science*, (4), Article 104. https://doi.org/10.1007/s41109-019-0228-y

Vermeer, S., Trilling, D., Kruikemeier, S., & de Vreese, C. (2020) Online news user journeys: The role of social media, news websites, and topics. *Digital Journalism*, *8*(9), 1114-1141. https://doi.org/10.1080/21670811.2020.1767509

Watanabe, K. (2013). The western perspective in Yahoo! News and Google News: Quantitative analysis of geographic coverage of online news. *International Communication Gazette*, *75*(2), 141-156. https://doi.org/10.1177/1748048512465546

Wilson, T. D., & Maceviciute, E. (2013). What's newsworthy about 'information seeking'? An analysis of Google's News Alerts. *Information Research*, *18*(1), Article 557. https://informationr.net/ir/18-1/paper557.html

Wilson, L. (2021). *How to get your website listed in Google News*. Search Engine Journal. https://www.searchenginejournal.com/how-to-get-listed-in-google-news/379701/

Wubben, S., van den Bosch, A., & Krahmer, E. (2011). Paraphrasing headlines by machine translation. Sentential paraphrase acquisition and generation using Google News. In T. Markus, P. Monachesi, & E. Westerhout (Eds.), *Computational Linguistics in the Netherlands 2010: Selected Papers from the Twentieth CLIN Meeting* (pp. 169-183). LOT. https://ilk.uvt.nl/~swubben/publications/clin_paraphrasing.pdf

Young, A., & Atkin, D. (2022). An agenda-setting test of Google News world reporting on foreign nations. *Electronic News, 17*(2), 113-132. https://doi.org/10.1177/19312431221106375

Young Lin, L. L., & Rosenkrantz, A. B. (2017). The U.S. online news coverage of mammography based on a Google News search. *Academic Radiology*, *24*(12), 1612-1615. https://doi.org/10.1016/j.acra.2017.05.011