



Peer-Reviewed Article

A Content Analysis of News Analyses: Examining Trends in News Content and Resources

Stacy Gilbert, *University of Colorado*

Rebecca Kelley, *Louisiana Board of Regents*

ABSTRACT

News websites and databases have changed over the last 20 years, yet little is known about the type of news content studied and how the ways researchers access content have evolved. This paper aims to identify trends in news analysis studies by examining 216 print and online news analyses published in journalism and mass communication studies journals by U.S. authors between 2002 and 2020. Each publication was coded for their methodological attributes. Findings show most studies analyze text articles. Subscription-based news aggregator databases like LexisNexis, NewsBank/Access World News, ProQuest, and Factiva are the most popular resources to access news content, and there has been a statistically significant increase in the use of news websites and public databases. Librarians can use these findings to assess their news collections and advise researchers on resources to access news content for news [research](#).

KEYWORDS

News databases, content sampling, newspapers, research methods, methodology trends

SUGGESTED CITATION

Gilbert, S.; & Kelley, R. (2024). Title of manuscript in sentence case. *Journal of New Librarianship*, 9(1), 1-25. <https://doi.org/10.33011/newlibs/15/1>

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Introduction

News content analyses are a popular area of study across disciplines (Feeney, 2014), and scholars who analyze news content frequently turn to library databases and librarians to carry out their studies. Online archives, databases, and websites, along with computer-aided text analysis (CATA) software and Application Programming Interfaces (APIs) have made it easier to search for and access news content (Krippendorff, 2004, Neuendorf, 2017), creating new possibilities for researchers' methodologies. News databases, websites, and online archives also presents new opportunities for academic librarians to support news research. Currently, for example, academic librarians advise on text and data mining (TDM) projects, yet emerging and evolving news databases and websites afford an opportunity for them to build their support for these and other projects. TDM support includes advising researchers about specific databases' content and search functions, programming, software, and other resources for scraping and analyzing news content, and any additional permissions or licenses unique to a particular resource (Anderson & Craiglow, 2017; Cheney, 2013; Wallace & Feeney, 2018).

Although many researchers study news content and libraries have the collections and expertise to support these projects, little is known about which types of messaging units (e.g., news articles, photos, readers' comments, etc.) are examined and how the use of news resources to access content have changed over time. As a result, librarians are not adequately informed about what units of news information need to be made easily discoverable, what resources need to be made particularly accessible, and what new tools and software need to be taught to researchers. The present study examines 18 years of print and online news content studies written by scholars in the United States (U.S.) from the mass communication and journalism disciplines which are known to frequently study news content (Riffe & Freitag, 1997). Through this study, academic librarians will have a better understanding of the types of news content used in academic studies and how news content was accessed by scholars over time, with broader implications of informing their news collections and consultation practices with researchers across disciplines.

Literature Review

Sampling with subscription news aggregator databases

In order to study news content, mass communication and journalism scholars first develop a sample of news content. Sampling involves developing a subset of content that is representative of the population, helping researchers economize their efforts (Krippendorff, 2004, p. 84). Prior research shows that U.S. scholars of the mass communication and journalism disciplines tend to develop their samples with content from major print legacy news outlets, newspapers that originated in print, like *The New York Times*, *The Washington Post*, or *USA Today* (Riffe & Freitag, 1997; Youngblood et al., 2013). A study of dissertations and theses

that used newspapers in the research found that the majority (67%) used only newspapers from the U.S., 23% used non-U.S. newspapers, and about 10% used both (Feeney, 2014).

Subscription news aggregator databases are frequently used by mass communication and journalism scholars to access news articles for analysis. These databases are appealing because users can access multiple newspapers and search thousands of articles with one search (Stryker et al., 2006). Academic libraries often subscribe to news aggregators like Nexis Uni (formerly LexisNexis Academic), Factiva, and NewsBank/Access World News. There is limited research on the most subscribed-to databases at academic libraries, but two surveys of business libraries' database collections found the leading news databases were LexisNexis products, Factiva, Newspaper Abstracts (from ProQuest), and NewsBank. Demonstrating how the popularity of databases can change over time, NewsBank did not appear in a 2003 survey of business libraries' database collections but was subscribed to by 49% of libraries over a decade later (Kim & Wyckoff, 2016; Schnedeker, 2003). The authors of the present study could only find two papers that analyzed the resources used to build samples for news analyses. First, Youngblood et al. (2013) found that from 2007 to 2011, the *Newspaper Research Journal* published 31 content analysis publications that used articles from *The New York Times* in their samples. Twenty of the 31 articles described how they access *The New York Times*: 18 of the 20 articles used LexisNexis Academic, and two used hard copy papers or microfilm. Eleven of the 31 articles did not provide clear or substantial information on how they accessed articles. These data have many limitations: it was presented as background information in an article, the authors only provided a brief description of their methodology, and their study looked at a small number of articles. Nonetheless, these findings provide an initial starting point for understanding how researchers access news articles.

A second study used EBSCO's Communication Source to find 346 content analyses of the news that were published from 2015 to 2020. The authors specifically looked at which databases the studies used, only examining articles that used Nexis Uni and other LexisNexis products, Factiva, Google News, ProQuest, and NewsBank. They found Nexis Uni and LexisNexis databases were used for 65.9% articles, followed by Factiva (18.2%), ProQuest (14.2%), Google News (13.3%), and NewsBank (5.5%). Most articles (83.5%) used a single database, but those articles that used more than one database often used Nexis Uni and LexisNexis products with ProQuest (Buntain et al., 2023). Both this study and Youngblood, et al. (2013) found LexisNexis Academic/Nexis Uni to be the most popular resource amongst communication and journalism scholars' news analyses.

News aggregator databases' search features and lack of certain content can present challenges to researchers. One limitation noted in the research is these databases offer various searching, indexing, and algorithm features that can retrieve different articles and complicate replicating studies' samples (Blatchford, 2020; Deacon, 2007; Riffe et al., 2019; Sabelhaus & Cawley, 2013; Woolley, 2000; Youngblood et al., 2013). News aggregator databases also have

incomplete coverage: not all news outlets and dates are available, nor are many articles from wire services, articles by freelancers, images and figures along with their captions, advertisements, and editorial letters (Deacon, 2007; Lacy et al., 2015; Orenstein, 1993; Ridout et al., 2012; Snider & Janda, 1998; Weaver & Bimber, 2008).

Furthermore, database design, features, and content change over time. For example, LexisNexis Academic was a news aggregator database to which many academic institutions subscribed. In 2017, LexisNexis launched a new database called Nexis Uni, designed for millennial users, and in 2020, Nexis Uni no longer included articles from *The Washington Post* (Beaujon, 2020; LexisNexis, 2017). Databases can also have exclusive rights to news outlets, forcing researchers to be selective over which publications or databases they include in their study. NewsBank is one example of a database with exclusive content—it has over 5,000 exclusive license titles (Ismail & Bareiss, 2021). Database vendors such as ProQuest may offer multiple news databases and various subscription models, causing researchers to lose or gain access to different news content at their own institution. Search settings controlled by individual libraries can vary between institutions, causing researchers to retrieve different results from the same search (Driedger & Weimer, 2015).

News on the Web

In addition to news aggregator databases, the advent of news websites, whose development rapidly grew in the mid 1990s to 2000s, offers new possibilities for researchers to study and access content. This includes the creation of digital-native or born-digital news websites, news outlets that originated on the web, such as Politico and BuzzFeed News. Legacy news organizations, outlets that did not originate on the web including print legacy or born-print papers (news media that originated in print) and television news networks, also launched their own websites. While there has been a proliferation of news websites in the last 20 years, print legacy papers have struggled in readership and revenue (Grundy, 2022), potentially limiting the number of print publications for researchers to study and newspaper databases to include in their institutions' collections.

The prevalence of news websites for researchers means that there is new content to study and new methods to access the news, fulfilling a gap in news aggregator databases' content which lacks born-digital news organizations (Gilbert & Watkins, 2020) and where coverage of smaller newspapers is inconsistent (Communication Studies Committee, 2022). News websites also have breaking news and developing content, reader comments, multimedia, and interactive stories that may not be part of subscription news aggregators' collection.

News websites, if allowed by publishers, can support web crawlers and APIs, making them ideal for text mining projects. Web portals (e.g., Yahoo!) and public aggregators (aggregators that do not require a subscription to search for content, e.g., Google News) make it easier for the public to discover and access tens of thousands of news websites globally, and they

link people to news articles and other news content (Salwen, 2004). Web portals and public aggregators are free to search but may require a subscription to news sites to access news content. Websites with historical news articles are also valuable resources for research projects. The most notable website might be *Chronicling America* which provides free access to news articles published between 1770-1963 and supports web crawlers and an API (Library of Congress, n.d.). News websites and public aggregators have limitations: users may need their own subscription to access news content and websites lack sophisticated search features often found in subscription news aggregators.

Research Purpose and Questions

Through collections and consultations, academic libraries play a role in how scholars discover and access news content. Understanding the type of news content scholars use and how they access content for research purposes can benefit librarians who advise on such projects. Data on what resources and methods were used, are currently in use, and are emerging can help librarians tailor consultations and workshops to the latest tools and methods. This paper builds upon previous studies that examined how researchers access news articles for content analysis studies and the news outlets they used to develop their samples (Buntain et al., 2023; Youngblood et al., 2013). By investigating 18 years of news analysis studies from communication, journalism, and media studies journals, three research questions are investigated:

RQ1: What are the characteristics of print and online news content (e.g., types of content, news outlets, etc.) used in mass communication and journalism's news analyses, and how have the characteristics changed over time?

RQ2: What resources do researchers in these disciplines use to access print and online news content, and how have the resources changed over time?

RQ3: Of all the resources used for research, which news aggregators do researchers use the most frequently, and how has that use changed over time?

Methodology

The authors conducted a content analysis of scholarly articles that examined news content from the mass communication and journalism fields. While numerous disciplines study the news, this paper focuses on news analyses published in communication, journalism, and media studies journals to create a manageable dataset that can be expanded on in follow up studies. These fields are an appropriate starting point for this search because they frequently conduct content analyses of the news (Riffe & Freitag, 1997), and the news industry is associated with these fields. To identify a sample, the authors used the EBSCO Communication & Mass Media Complete (CMMC) database because it could be easily accessed by the authors whose

institutions subscribe to it, it covers key journalism and communication journals, and it has a thesaurus to select relevant subject terms.

The search is limited to scholarly articles published from 2000 to 2021, although the final date range used for analysis is 2002 to 2020 as noted in the Results section below. This time span covers changes in the journalistic landscape, such as Google News' launch in 2002, the development of legacy news' websites, and the expansion of digital native news. It also includes a period before and after content analysis grew significantly across disciplines in the mid-2000s (Neuendorf, 2017) and shifts in academic news aggregator databases occurred (e.g., LexisNexis Academic becoming Nexis Uni).

The authors searched CMMC with the following criteria:

- Subject terms:
 - “content analysis” OR “rhetorical analysis” OR “discourse analysis” OR “thematic analysis” OR “sentiment analysis”
 - AND newspapers OR “electronic newspapers” OR “news websites” OR “news aggregators”
- Date range: January 1, 2000 – April 30, 2021
- Scholarly (Peer Reviewed Journals) only
- Language: English
- Source types: Academic Journals

This search was conducted on May 7, 2021, and yielded 537 results. The authors of the present study, both librarians in the U.S., went through the results and excluded additional articles based on the following criteria, reducing the sample to 218 articles:

- Excluded articles that did not have one author associated with a U.S. institution, since the subscription-based news aggregator databases available at most U.S. academic institutions may not have been available to international scholars.
- Excluded articles when the entire article was not available.
- Only included articles that examined newspapers or news websites. Included broadcast, cable, or radio news' websites, but excluded articles that looked at TV or radio broadcasts.

Content Analysis

Content analysis involves systematic coding of variables to find patterns in the sample. The authors first determined what types of variables were critical to answering the research questions and wrote codes around these variables. They practiced coding together and later

individually on a small selection of random articles and continued to refine the codes and code definitions. As the researchers became immersed in the sample, new codes emerged and were added to the codebook. The final codebook has six categories that analyzed each scholarly article's methodology. The categories and corresponding codes are in Appendix 1. The six categories are called:

1. Year(s) news articles in the scholarly papers' samples were published
2. Number of years between publication of the news articles and the scholarly paper
3. Number of news outlets in the scholarly papers' samples, such as if sample was a predetermined set list of news outlets or if the scholarly articles used all available news outlets in a resource.
4. Types of news outlets (e.g., non-U.S. legacy newspaper, U.S. TV news website, etc.) in the scholarly papers' samples
5. Types of news content (e.g., text-based content, images, reader comments, etc.) in the scholarly papers' samples
6. Resources scholarly papers used to access the samples, such as subscription-based news aggregator databases, microfilm, websites, etc.

Intercoder reliability (ICR) was conducted to measure the coders agreement of the codebook. Both authors coded a simple random selection of seventy-two out of the 218 articles, or 33.0% of the sample. The 72 articles meet the sample size requirements that Lacy and Riffe (1996) determined for a legitimate ICR. Dedoose, a qualitative and mixed-methods research tool, was used to calculate Cohen's kappa for ICR. Twenty-five codes had a kappa coefficient of .81 to 1.00, while six codes had a kappa coefficient between .61 to .80. According to Landis and Koch (1977 as cited in Neuendorf, 2017, p. 168), kappa coefficients between .81 to 1.00 are considered to be "almost perfect" agreement, while kappa coefficients between .61 to .80 are in substantial agreement. Cohen's kappa is reported for all the codes in Appendix 1.

A total of 12 codes were not measured for ICR. Of these twelve, eight codes were applied zero or one time in the 72-article sample set, and therefore were not applied enough to be calculated. The ICR for another four codes under "Date range" that measured the length of time since the scholarly article was published compared to the scholarly article's sample date could not be calculated due to limitations in Dedoose's ICR test platform.

Data Analysis

The data was analyzed by percentage of all articles that contained a given independent variable (e.g., types of news outlets examined, number of news outlets examined, resources used to access samples, etc.). The percentage was calculated and analyzed by year, four or five-year strata (e.g., 2002-2006, 2007-2011, 2012-2016, and 2017-2020), and the study's time period.

To analyze the trend across time, the percentage of articles that contained a given independent variable was calculated for each year. The percentages per year were analyzed using ordinary least squares (OLS) simple linear regressions of the dependent variable on year. As there was only one paper published in both 2002 and 2003, the percentage of papers with a given variable for the years would be 100% and skew the linear regression. Thus, linear regression analysis omitted the years 2002 and 2003 and the linear regression looks at the 2004 to 2020 period. The goodness of fit for the regression equation is reported using R^2 . For interpreting the p -value, the authors used an alpha level of .05 for all statistical tests. P -values less than .05 means that the correlation is statistically significant. Further qualitative analysis was done informally by reviewing how the codes manifested in the scholarly papers' texts.

Results

The authors coded the 218 scholarly articles that met their specified criteria. After coding was completed, the authors excluded two 2021 publications from the study, since they did not have a complete year of data, and this would have skewed the results which were analyzed by percent published per year. Additionally, 2000 and 2001 were excluded from the study as the sample did not include any publications from those years. Therefore, the study's sample consisted of 216 scholarly articles published between 2002 and 2020. Table 1 outlines the total number and overall percentage of all scholarly publications used in this study by methodological attribute.

Table 1

Methodological Attributes of Scholarly Papers' Content Analyses

Code	Variables	%	(<i>n</i> = 216)
Number of years since the news articles in samples were published and the scholarly papers were published	0-9 years ago	91.2%	197
	10-19 years ago	25.0%	54
	20+ years ago	14.4%	31
	Did not disclose dates	2.3%	5
Types of news content in the scholarly papers' samples	Text based	92.1%	199
	Other	10.7%	23
	Photographs	9.3%	20
	Reader Comments	2.3%	5
Types of news outlets in the scholarly papers' samples	Legacy newspapers (U.S.)	74.1%	160
	Legacy newspapers (non-U.S.)	18.9%	41

	Legacy news websites (U.S.)	10.2%	22
	TV news websites (U.S.)	7.9%	17
	Digital native news websites	6.5%	14
	Legacy news websites (non-U.S.)	3.7%	8
	TV news websites (non-U.S.)	1.4%	3
Number of news outlets in the scholarly papers' samples	Used specific news outlets	79.2%	171
	1 outlet	13.9%*	30
	2 outlets	13.9%*	30
	3-5 outlets	25.5%*	55
	6+ outlets	25.9%*	56
	Used all news outlets in database	21.3%	46
Resources scholarly used to access the samples	Subscription-based news aggregator databases	56.9%	123
	LexisNexis Academic	36.6%*	79
	Factiva	3.2%*	7
	ProQuest	2.8%*	6
	Other subscription-based news databases	14.4%*	31
	Nexis Uni	0%	0
	Did not disclose resources used	25.9%	56
	News websites	19.9%	43
	Other public databases	5.6%	12
	Hard copies of newspapers	5.1%	11
	Microfilm/microfiche	3.7%	8
	Public search engine	1.4%	3
	Clipping service	0.9%	2
	Historical newspaper collection	0.5%	1

Note. Percentages may not add up to 100 because some scholarly papers were included in multiple categories. *Percentage of all scholarly papers in the sample.

Characteristics of News Articles

RQ1 asks what kinds of print and online news content researchers use in news analyses and how these characteristics change over time. Over 90% of scholarly papers (91.2%, $n=197$) analyzed news stories that were published zero to nine years prior to the scholarly papers' publication (see Table 1), which is important for understanding that most studies analyze contemporary news. Ninety-two percent (92.1%, $n=199$) of scholarly publications studied text-

based news content in their samples. Few studies analyzed photographs (9.3%, $n=20$), reader comments (2.3%, $n=5$), and other content (10.7%, $n=23$).

The following types of news outlets and platforms were found to be used in the scholars' studies:

- U.S. or non-U.S. legacy newspapers: born-print newspapers, including legacy news outlets that were purchased by TV or digital media publishers.
- U.S. or non-U.S. legacy news websites: born-print newspapers websites or web applications, including online-only content from legacy news websites.
- U.S. or non-U.S. TV news websites: cable or local TV news websites or web applications.
- U.S. or non-U.S. digital native news websites: digital native news websites or web applications. The term "websites" was defined by the scholarly papers' authors, and sometimes included search engines (e.g., Yahoo!, Google News) if these met the authors' definition.

Legacy newspapers are included in a majority of news analyses (see Table 1). About three-fourths of the scholarly papers had samples comprised of U.S. legacy newspapers news outlets (74.1%, $n=160$), followed by non-U.S. legacy newspapers (18.9%, $n=41$). U.S. legacy news websites, (10.2%, $n=22$), U.S. TV news websites (7.9%, $n=17$), and digital native news websites (6.5%, $n=14$) were also used in the samples. While the majority of researchers used U.S. legacy newspapers in their samples, this is decreasing over time. From 2002-2006 period, 78% of articles used U.S. legacy newspapers, while in 2017-2020, this number decreased to 69% (see Table 2). Figure 1 shows, from 2004-2020, a statistically significant decrease in news analyses that use U.S. legacy newspapers ($R^2=0.24$, $p=0.04$). At the same time, non-U.S. legacy news websites ($R^2=0.24$, $p=0.04$), U.S. digital native news websites ($R^2=0.27$, $p=0.03$), U.S. TV news websites ($R^2=0.51$, $p=0.001$), and non-U.S. TV news websites ($R^2=0.25$, $p=0.04$) saw a statistically significant increase from 2004 to 2020. As demonstrated in Table 2, none of the articles published between 2002-2006 studied U.S. TV news websites, but this increased to 11% of articles published between 2017-2020. Also notable is 6% of articles studied U.S. legacy news websites from 2002 to 2006, but this increased to 25% of articles studying these outlets in 2012 to 2016 and 16% in 2017 to 2020. Non-U.S. legacy newspapers ($R^2=0.05$, $p=0.37$) and U.S. legacy news websites ($R^2=0.12$, $p=0.18$) also increased during this time but it was not statistically significant (see Figure 1). In summary, from 2002 to 2020, the majority of papers used U.S. legacy newspapers, but the percent of articles studying these sources decreased over time. U.S. digital native news websites and U.S. legacy news websites saw some of the biggest increases.

Table 2

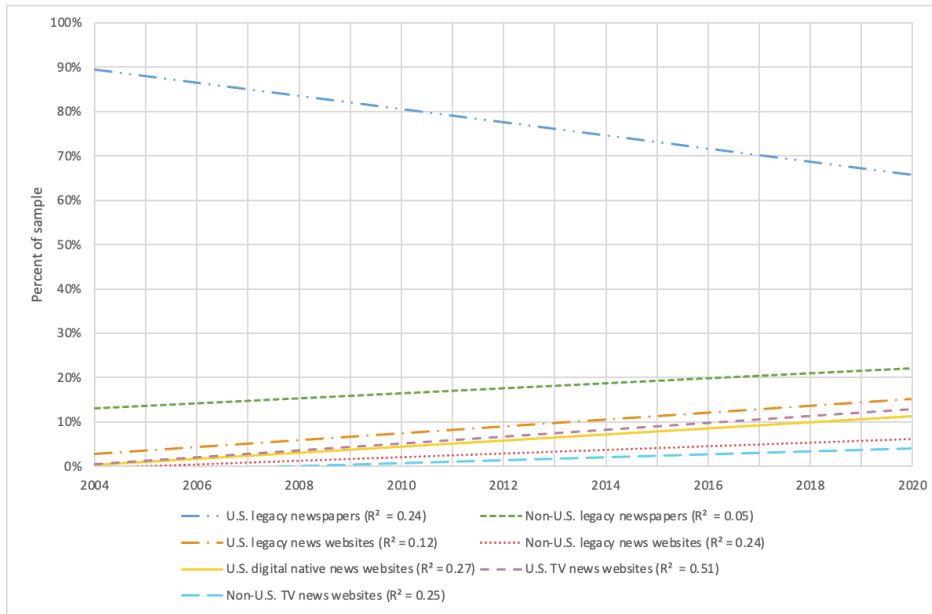
Percentage of Articles Within 5 or 4-Year Strata That Used Specific Types of News Outlets, 2002-2020; Data Presented as % (n)

Year	U.S. legacy newspapers	Non-U.S. legacy newspapers	U.S. legacy news websites	Non-U.S. legacy news websites	U.S. digital native news websites	U.S. TV news websites	Non-U.S. TV news websites
2002-2006 (n = 18)	78% (14)	17% (3)	6% (1)	0% (0)	6% (1)	0% (0)	0% (0)
2007-2011 (n = 69)	83% (57)	17% (12)	10% (7)	1% (1)	4% (3)	7% (5)	0% (0)
2012-2016 (n = 65)	69% (45)	25% (16)	6% (4)	3% (2)	5% (3)	8% (5)	2% (1)
2017-2020 (n = 64)	69% (44)	16% (10)	16% (10)	8% (5)	11% (7)	11% (7)	3% (2)

Note. Percentages may not add up to 100% because some articles may have used more than one type of news outlet.

Figure 1

Types of News Outlets in the Scholarly Papers' Samples, 2004-2020, Linear Trend

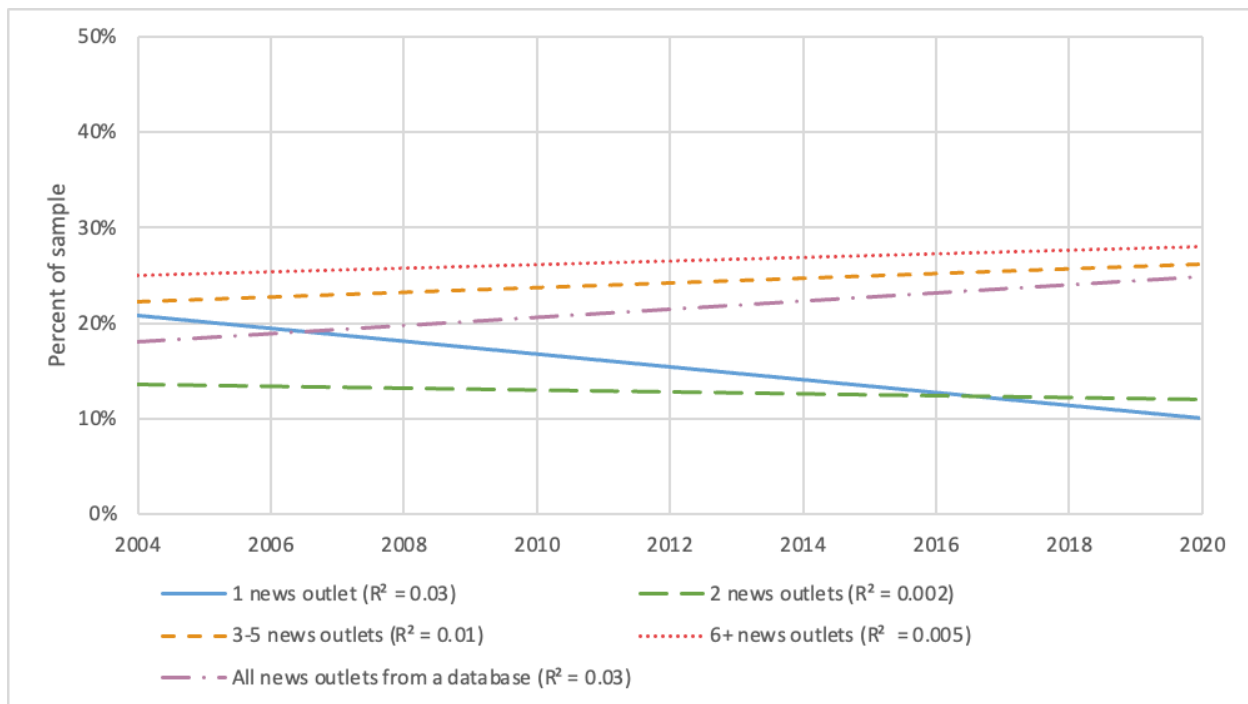


Alt text: Linear chart of types of news outlets analyzed by scholarly papers from 2004 to 2020 where U.S. legacy newspapers is used by majority of studies but is decreasing. Other types of news outlets are seldom used but are increasing.

Most scholarly papers analyzed units from specific news outlets (see Table 1). Overall, 79.2% ($n=171$) used a list of pre-determined news outlets, while 21.3% ($n=46$) of scholarly publications' samples included any outlets in a news database that met their search criteria. When researchers analyzed content from specific news outlets, 25.9% ($n=56$) of all scholarly publications sampled from six or more news outlets and another 25.5% ($n=55$) sampled from three to five news outlets. The remaining scholarly papers limited their sampling to either two news outlets (13.9%, $n=30$) or one news outlet (13.9%, $n=30$). As demonstrated in Figure 2, there is very little change over time in the number of news outlets used in scholar papers' samples as there were no statistically significant trends.

Figure 2

Number of News Outlets in the Scholarly Papers' Samples, 2004-2020, Linear Trend



Alt text: Linear chart of the number of news outlets used in the news analyses from 2004-2020 where all number of news outlets are between about 10% to 30% and use of one news outlet is decreasing.

The authors informally analyzed the scholarly publications that were sampled from one or two news outlets. Of these 60 publications, almost half of them ($n=29$) analyzed content from *The New York Times*. Thirty publications specifically sampled from international newspapers, many of which were from Chinese newspapers. Many of the scholarly publications with only one or two outlets sampled from smaller, local, or regional newspapers.

Resources to Access News Content

RQ2 asks what resources researchers use to access the print and online news content and how use of these resources changes over time. Scholars relied the most on subscription-based news aggregator databases. About half (56.9%, $n=123$) of the scholarly publications accessed their samples from subscription-based news aggregator databases (see Table 1). Additionally, as Figure 4 demonstrates, the percentage of the sample that accessed news content from subscription-based news databases trended upward slightly, although this is not statistically significant ($R^2=0.08$, $p=0.27$). About half (52%) of the articles published between 2017 to 2020 used a subscription-based news database (see Table 3).

Table 3

Top Three Types of Resources Articles Used to Access News Content, by Percentage of Articles Within 5 or 4-Year Strata, 2002-2020; Data Presented as % (n)

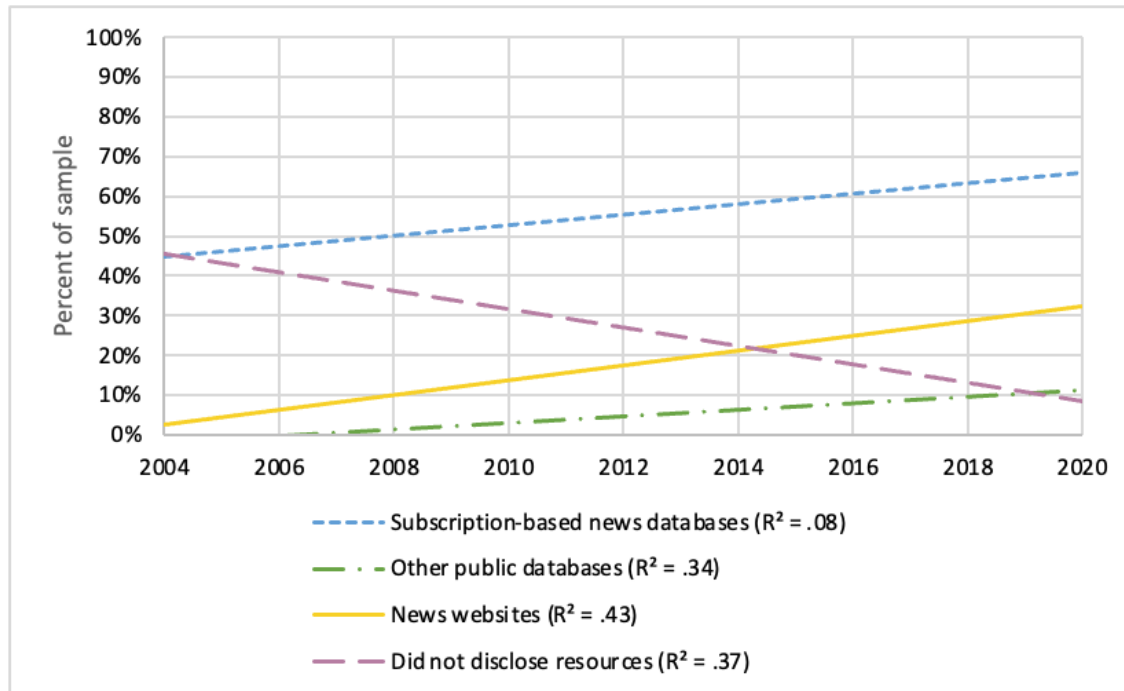
Year	Other public data-bases	News website	Did not disclose resource	All subscription-based news databases	All subscription-based news databases			
					Lexis Nexis	Factiva	Pro Quest	Other subscription-based databases
2002-2006 ($n=18$)	0% (0)	0% (0)	44% (8)	28% (5)	22% (4)	0% (0)	0% (0)	6% (1)
2007-2011 ($n=69$)	4% (3)	4% (3)	39% (27)	52% (36)	41% (28)	0% (0)	3% (2)	9% (6)
2012-2016 ($n=65$)	3% (2)	3% (2)	18% (12)	75% (49)	45% (29)	8% (5)	3% (2)	20% (13)
2017-2020 ($n=64$)	11% (7)	11% (7)	14% (9)	52% (33)	28% (18)	3% (2)	3% (2)	17% (11)

Note. Some articles may have used more than one resource.

The second most used resource was news websites, which were used by 19.9% ($n=43$) of researchers (see Table 1). This included websites from legacy newspapers, digital native news, and TV news. News websites first appeared in the sample in 2005 and became more commonly included by 2008. Regression analysis shows news websites had a statistically significant increase in use ($R^2=0.43$, $p < .001$) as shown in Figure 3.

Figure 3

Top Three Types of Resources Scholarly Papers Used to Access Their Samples, 2004-2020, Linear Trend



Alt text: Linear chart of percent of papers that used subscription-based news databases, other public databases, and news websites, as well as did not disclose resources from 2004-2020 with subscription-based news databases used the most and is increasing.

The third most used resource was “other public databases,” which was used by only 5.6% ($n=12$) of articles in the sample (see Table 1). “Other public databases” includes “Today’s Front Pages” archive at Newseum.org, the Internet Archive’s Wayback Machine, Campaign Mapping Project, and other publicly-available online databases of news content. However, as Figure 3 shows, there was a statistically significant increase in scholarly publications using public databases to access their samples across the 18-year period ($R^2=0.34$, $p=0.01$).

A small percentage of articles used other resources. This includes hard copies of newspapers (5.1%, $n=11$); microfilm or microfiche (3.7%, $n=8$); public search engines like Google, Google News, or Google Site Search (1.4%, $n=3$); clipping services (0.9%, $n=2$); and historical newspaper collections like *Chronicling America* (0.5%, $n=1$). In summary, subscription-based news aggregator databases were used the most frequently (56.9%, $n=123$), followed by news websites (19.9%, $n=43$). News websites are also seeing increased use over time.

Of note, 25.9% ($n=56$) of all scholarly publications in the sample either did not disclose the resources used to access their sample or did not provide enough information for the authors

of this paper to categorize the resources used (see Table 1). However, as seen in Figure 3, when analyzing the percentage of scholarly publications in the sample over time, there was a statistically significant downward trend in scholarly papers that did not disclose the resources used to access their samples ($R^2=0.37$, $p=0.01$).

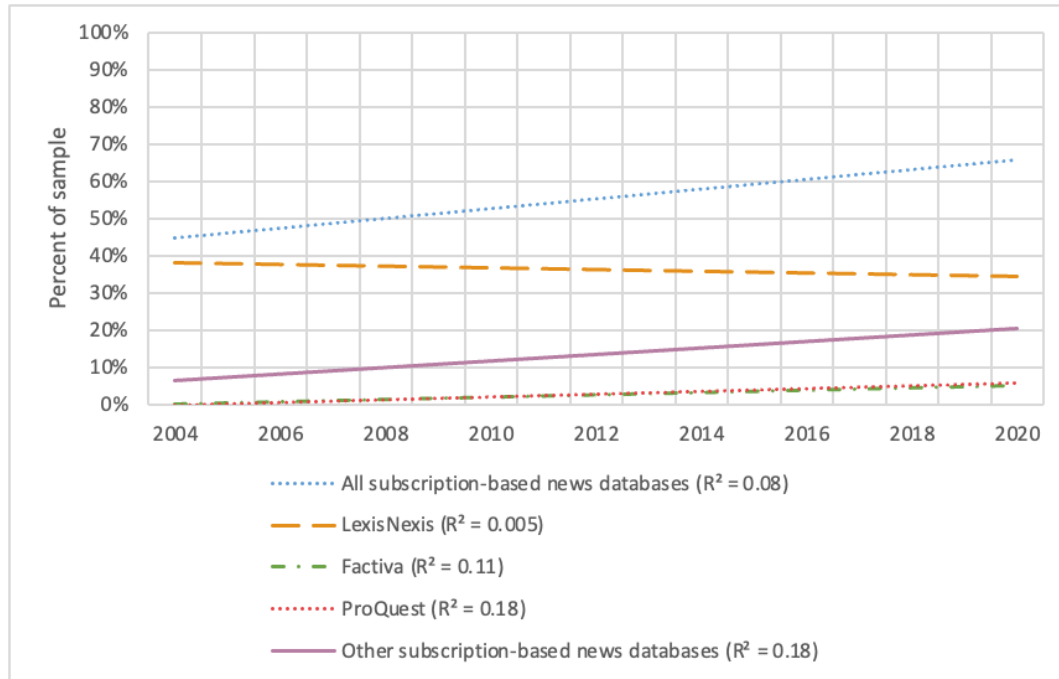
Subscription-Based News Aggregator Databases

Q3 asks which news aggregator databases do researchers use and how has use of these resources changed over time. As previously mentioned, 56.9% ($n=123$) of the sample accessed news content with subscription-based news aggregator databases. The authors coded the samples for some commonly used subscription-based news aggregator databases, including LexisNexis Academic, Nexis Uni, Factiva, and ProQuest. The authors originally included NewsBank/Access World News in the codebook, but it was removed when it did not show up in the testing stage. About one-third (36.6%, $n=79$) of all scholarly publications reported using LexisNexis Academic to access their sample (see Table 1). Factiva used in 3.2% ($n=7$) of articles, while ProQuest news databases were used in six scholarly publications (2.8%). The remaining subscription-based news databases were not coded individually, yet they accounted for 14.4% ($n=31$) of all scholarly publications in the sample. An informal review of the publications that were coded as using “other subscription-based databases” revealed that a third of these publications used NewsBank/Access World News ($n=10$) to access their sample of news content. “Other subscription-based databases” used by researchers included unspecified EBSCO databases, unspecified “library databases,” and international or foreign-language news databases.

The trendline in Figure 4 shows a slight decrease in the use of LexisNexis Academic ($R^2=0.005$, $p=0.79$), but this is not statistically significant. As seen in Table 3, over the 18-year period studied, the percentage of articles using LexisNexis increased from 22% of articles published between 2002 to 2006, to 45% of articles published between 2012 to 2016. It then decreased to 28% of articles published between 2017 to 2020. The use of Factiva ($R^2=0.11$, $p=.20$), ProQuest ($R^2=0.18$, $p=0.08$), and “other subscription-based news databases” (e.g., NewsBank/Access World News) ($R^2=0.18$, $p=0.09$) increased over time, but are not statistically significant. Factiva was not used in scholarly publications until 2012, and ProQuest news databases were first used in 2010. According to Table 3, Factiva and ProQuest have each seen steady use in more recent years. Articles that used other subscription-based databases increased from 6% in 2002-2006 to 20% in 2012-2016 and then 17% in 2017-2020.

Figure 4

Subscription-Based News Aggregator Databases Scholarly Papers Used to Access Their Samples, 2004-2020, Linear Trend



Alt text: Linear chart of percent of subscription-based news aggregator databases used in news analyses from 2004-2020 with LexisNexis used the most at 30-40% and other subscription-based news databases seeing increased use.

None of the scholarly publications in the study named Nexis Uni as a resource even though LexisNexis Academic was replaced with Nexis Uni in 2017. However, an article published in 2019 and another published in 2020 noted the use of LexisNexis Academic in their methodology. The researchers from the 2019 publication accessed their sample in 2017, but the researchers in the 2020 publication retrieved data in 2020 and possibly used a different LexisNexis product. Informal analysis of the data from 2018 to 2020, show the leading resources researchers used were news websites ($n=10$), “other subscription-based databases” ($n=7$), and “other public databases” ($n=4$).

Discussion

In order for academic librarians to support researchers in finding and accessing news content, they should be familiar with the latest developments in scholarly news research. This paper analyzed scholarly articles from communication, journalism, and media studies journals to better understand what kind of news content and news outlets researchers from these fields

study, and what kind of resources they use to access news content. While research is focused on select disciplines, librarians who support other researchers may use the findings to understand the news research needs of other scholars.

The present study found communication and journalism news analyses published between 2002 and 2020 primarily study news content created within nine years of the articles' publication (91.2%, $n=197$) and text-based content (92.1%, $n=199$). Most scholarly papers studied legacy newspapers from the U.S., but there is a statistically significant increase in the use of non-U.S. legacy newspapers, non-U.S. news websites, U.S. digital native news websites, U.S. TV news websites, and non-U.S. TV news websites. The increased interest in online news mirrors the proliferation of these outlets that started in the mid-1990s and continued in the 2000s (Salwen, 2004). Academic librarians can use this finding to assess their collections and the discovery pathways on their websites to accommodate the growing interest in U.S. and foreign titles and news websites that support scholars' research.

Online archives and news aggregator databases provide researchers access to a large quantity of news content (Neuendorf, 2017; Stryker et al., 2006). However, this study found 79.2% ($n=171$) of scholarly articles focused their research on specific news outlets instead of all the content available in a database, and there were no statistically significant trends in the number of news outlets used. While news aggregators or websites provide researchers with the option to study hundreds of media outlets, this finding indicates that scholars' studies are selective. Informal analysis shows *The New York Times*, Chinese publications, and local newspapers were specifically included in news studies. Further research could analyze if scholars in other disciplines also focus on specific news outlets, and if so, which publications. In the meantime, librarians can use these findings to assess their collections and if their websites offer ways for scholars to access specific news outlets, particularly *The New York Times*, Chinese news media, and local news.

This study corroborated prior research in finding news aggregator databases are the primary resources that research use to access news articles, and that LexisNexis Academic is a popular resource among researchers (Buntain et al., 2023; Youngblood et al., 2013). LexisNexis was widely available at academic business libraries in the U.S. and its availability may have contributed to its use in research projects (Kim & Wyckoff, 2016; Schnedeker, 2003). The present study analyzed Nexis Uni and LexisNexis Academic separately, finding 36.6% ($n=79$) of scholarly papers used LexisNexis Academic, but no studies reported using their new database, Nexis Uni. It is possible researchers are using Nexis Uni but are confused about the name or have not internalized the name change. Including "LexisNexis Academic" to Nexis Uni library records or research guides may help researchers discover this database.

Like Buntain and colleagues (2023), this study also found researchers use the databases ProQuest, Factiva, and NewsBank, although there was not a statistically significant increase

from 2004 to 2020. NewsBank/Access World News Research Collection increased use ($n=10$, categorized under “other subscription-based news databases”) mirrors the database’s growth in the number of publication titles and academic business library subscriptions over the last two decades (Gold, 2002; Ismail & Bareiss, 2021; Kim & Wyckoff, 2016; Schnedeker, 2003). Academic libraries have also been long-time subscribers to Factiva and ProQuest (Kim & Wyckoff, 2016; Schnedeker, 2003), but these databases were used less than LexisNexis. This data can give librarians a datapoint when analyzing news aggregator database for collections. Librarians may want to focus their attention on learning about these resources’ features and content and highlight them in research guides, consultations, and workshops.

Over time there was a statistically significant increase with scholars using news websites (19.9%, $n=43$), and “other public databases” (5.6%, $n=12$) like the Wayback Machine and the Campaign Mapping Project. However, one type of resource that is getting little use among scholars is search engines (e.g., Google News, Google, etc.). Although Google News is one of the top websites focused on the news in the U.S. and accessible by most people (Majid, 2022), this paper found only three papers (1.4%) used a search engine to access their samples. In comparison, Buntain and colleagues (2023) found that out of 346 studies, 13.3% used Google News. This discrepancy warrants further study. The increased use of news websites and public databases could be driven by the ability of most people to access them or purchase their own subscriptions, the rise of news websites (Salwen, 2004) coupled with the struggling print legacy papers (Grundy, 2022), the lack of born digital news content in subscription-based news aggregators (Gilbert & Watkins, 2020), or the development of computerized tools that assist in content analysis (Neuendorf, 2017). As scholars in the mass communication and journalism discipline are finding news in alternative resources to subscription-based news aggregators, librarians may want to familiarize themselves with news websites and public databases’ affordances and limitations (e.g., broken links, paywalls, missing content, and inadequate search features) and teach or develop guides to using these resources and tools. Further research is needed to analyze if access and discovery to a database or website drives use.

Despite the advances in news databases and websites, a small number of researchers continue to use physical formats for accessing the news. Hardcopies of newspapers were used by 5.1% ($n=11$) of the scholarly papers and microfilm/microfiche were used by 3.7% ($n=8$) of the scholarly papers. Additional research into the use of hardcopies would be helpful for understanding scholars’ perceived benefits of hardcopies and limitations of other news formats.

It is possible the resources used to access content for the sample and the characteristics of the news contents’ sample (e.g., types of news outlets, date range, and the content analyzed) are related. This study found authors note which news outlets are excluded from databases and need to use more than one resource to access their samples. If researchers are committed to including a specific news outlet, then they must use a resource that has access to that content. Given the findings that many scholarly articles studied *The New York Times*, this could

potentially explain why several scholarly articles used LexisNexis Academic, Factiva, or ProQuest as they provide access to this publication. If researchers are committed to using a specific database their institution subscribes to, then they could be limited in which news outlets and news content they can study. As news aggregator databases usually contain text-only content (Orenstein, 1993), this could explain why most articles ($n=199$) studied news' text rather than photos or other content. How researchers decide which messaging units to include in a sample and the resources used to access the messages warrants further study.

Conclusion and Limitations

Academic libraries support news research through collections, instruction, and workshops, and thus it is helpful to understand the type of news content researchers study and how they access it. This paper investigated trends in print and online news content analysis studies by conducting a content analysis of 216 papers published between 2002 and 2020 from the mass communication and journalism discipline. Overall findings suggest that three or more media outlets are being used in news content analysis studies. Subscription-based news aggregator databases are frequently used to access the news content sample, while news websites and public databases are also seeing increased use. Although LexisNexis Academic was once a popular resource to access sample news stories, its new database, Nexis Uni, was not named as a resource. Other resources may be taking its place: ProQuest, news outlets' websites, and public databases with news content have trended upward.

This research has some limitations. The paper focused on print and online news, excluding formats like television and radio transcripts. The database Communication & Mass Media Complete (CMMC) from EBSCO was used to find the sample of scholarly papers. This database has some of the leading communication, mass media, and journalism journals; however, some journals have an embargo period where their articles are not readily available through CMMC. CMMC also does not have a key journalism journal, *Digital Journalism*. Search terms used to build the scholarly article sample tend to be used for traditional news analysis studies, overlooking terms that would have included big data or crawling studies. Thus, this study is likely missing relevant scholarly papers and underreported larger news content studies. The sample was limited to authors affiliated with U.S. institutions. A broader sample would have provided additional publications that would have been of value.

Despite these limitations, this research can provide librarians with an understanding of how news analysis research has evolved to inform what types of news content need to be made discoverable and what resources and tools need to be taught and made accessible. Future research could explore news content analyses from other disciplines, why researchers select specific resources for accessing their samples, what resources are used in big data or crawling studies, how they decide which news outlets to include, and if researchers are aware of the limitations of news aggregator databases.

References

- Anderson, C. B., & Craiglow, H. A. (2017). Text mining in business libraries. *Journal of Business & Finance Librarianship*, 22(2), 149–165.
<https://doi.org/10.1080/08963568.2017.1285749>
- Beaujon, A. (2020, August 10). The Washington Post will soon be gone from LexisNexis. *Washingtonian*. <https://www.washingtonian.com/2020/08/10/the-washington-post-will-soon-be-gone-from-lexisnexis/>
- Blatchford, A. (2020). Searching for online news content: The challenges and decisions. *Communication Research and Practice*, 6(2), 143–156.
<https://doi.org/10.1080/22041451.2019.1676864>
- Buntain, N., Liebler, C. M., & Webster, K. (2023). Database use, database discrepancies: Implications for content analyses of news. *Newspaper Research Journal*, 1–16.
<https://doi.org/10.1177/07395329231155193>
- Caudle, D. M., Schmitz, C. M., & Weisbrod, E. J. (2013). Microform—Not extinct yet: Results of a long-term microform use study in the digital age. *Library Collections, Acquisitions, and Technical Services*, 37(1–2), 2–12. <https://doi.org/10.1016/j.lcats.2013.02.001>
- Cheney, D. (2013). *Text mining newspapers and news content: New trends and research methodologies*. 1–5. <https://library.ifla.org/id/eprint/233>
- Cheney, D., Knapp, J., Alan, R., & Czaplá, P. (2006). Convergence in the library's newsroom: Enhancing news collections and services in academic libraries. *College & Research Libraries*, 67(5), 395–417. <https://doi.org/10.5860/crl.67.5.395>
- Communication Studies Committee. (2022, October 26). *Newspaper Aggregators*. ACRL/EBSS Library Resources for Communication Studies (LRCS).
<https://acrl.libguides.com/ebss/lrcs/newspaperdatabases>
- Deacon, D. (2007). Yesterday's papers and today's technology: Digital newspaper archives and 'push button' content analysis. *European Journal of Communication*, 22(1), 5–25.
<https://doi.org/10.1177/0267323107073743>
- Driedger, S. M., & Weimer, J. (2015). Factiva and Canadian Newsstand Major Dailies: Comparing retrieval reliability between academic institutions. *Online Information Review*, 39(3), 346–359. <https://doi.org/10.1108/OIR-11-2014-0276>
- Feeney, M. (2014). Understanding news researchers through a content analysis of dissertations and theses. *Qualitative and Quantitative Methods in Libraries*, 3(1), 263–270.

- Freeland, M., & Bailey, M. (2008). Print newspapers: Are they still being used in academic and research libraries? *The Serials Librarian*, 55(1–2), 210–226.
<https://doi.org/10.1080/03615260801970881>
- Gilbert, S., & Watkins, A. (2020). A comparison of news databases' coverage of digital-native news. *Newspaper Research Journal*, 41(3), 317–332.
<https://doi.org/10.1177/0739532920950039>
- Grundy, A. (2022, June 7). *Service annual survey shows continuing decline in print publishing revenue*. United States Census Bureau. <https://www.census.gov/library/stories/2022/06/internet-crushes-traditional-media.html>
- Gold, H. (2002). NewsBank. *The Charleston Advisor*, 4(1), 48–48.
- Ignatow, G. (2023). *Text Mining*. <https://doi.org/10.4135/9781526421036821928>
- Ismail, L., & Bareiss, W. (2021). Newsbank: Access to World News Research Collection. *The Charleston Advisor*, 22(4), 49–55. <https://doi.org/10.5260/chara.22.4.49>
- Kim, K., & Wyckoff, T. (2016). What's in your list?: A survey of business database holdings and funding sources at top academic institutions. *Journal of Business & Finance Librarianship*, 21(2), 135–151. <https://doi.org/10.1080/08963568.2016.1140548>
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed). Sage.
- Lacy, S., & Riffe, D. (1996). Sampling error and selecting intercoder reliability samples for nominal content categories. *Journalism & Mass Communication Quarterly*, 73(4), 963–973. <https://doi.org/10.1177/107769909607300414>
- Lacy, S., Watson, B. R., Riffe, D., & Lovejoy, J. (2015). Issues and best practices in content analysis. *Journalism & Mass Communication Quarterly*, 92(4), 791–811.
<https://doi.org/10.1177/1077699015607338>
- LexisNexis. (2017, January 19). *LexisNexis Launches New Academic Research Solution Designed for Millennials*. LexisNexis. <https://www.lexisnexis.com/en-us/about-us/media/press-release.page?id=1484689760846412&y=2017>
- Library of Congress. (n.d.). *Chronicling America*. Retrieved September 9, 2022, from <https://chroniclingamerica.loc.gov/>
- Majid, A. (2022, December 13). *Top 50 news sites in the US in November: New York Times fastest-growing followed by MSN and BBC*. Press Gazette.
https://pressgazette.co.uk/media-audience-and-business-data/media_metrics/most-popular-websites-news-us-monthly-3/
- Neuendorf, K. A. (2017). *The content analysis guidebook* (Second). SAGE Publications Inc.

- Orenstein, R. M. (1993). "How full is full" revisited: A status report on search full-text periodicals. *Database*, 16(5), 14–23.
- Ridout, T. N., Fowler, E. F., & Searles, K. (2012). Exploring the validity of electronic newspaper databases. *International Journal of Social Research Methodology*, 15(6), 451–466. <https://doi.org/10.1080/13645579.2011.638221>
- Riffe, D., & Freitag, A. (1997). A Content analysis of content analyses: Twenty-five years of Journalism Quarterly. *Journalism & Mass Communication Quarterly*, 74(3), 515–524. <https://doi.org/10.1177/107769909707400306>
- Riffe, D., Lacy, S., Watson, B. R., & Fico, F. (2019). *Analyzing media messages: Using quantitative content analysis in research* (Fourth edition). Routledge, Taylor & Francis Group.
- Roberts, J. (2022). Newspapers and the persisting usefulness of microfilm: A matter of context. *Music Reference Services Quarterly*, 25(1), 19–31. <https://doi.org/10.1080/10588167.2021.1991165>
- Sabelhaus, L., & Cawley, M. (2013). Searching for news online: Challenging traditional methods. *Online Searcher*, 37(2), 10–14.
- Salwen, M. B. (2004). Online news trends. In *Online News and the Public* (1st ed., pp. 47–79). Routledge.
- Schnedeker, D. (2003). Business and economic databases access in academic business libraries. *Journal of Business & Finance Librarianship*, 9(1), 37–47. https://doi.org/10.1300/J109v09n01_04
- Snider, J. H., & Janda, K. (1998). *Newspapers in Bytes and Bits: Limitations of Electronic Databases for Content Analysis*. 29. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.502.7296&rep=rep1&type=pdf>
- Stryker, J. E., Wray, R. J., Hornik, R. C., & Yanovitzky, I. (2006). Validation of database search terms for content analysis: The case of cancer news coverage. *Journalism & Mass Communication Quarterly*, 83(2), 413–430. <https://doi.org/10.1177/107769900608300212>
- Wallace, N., & Feeney, M. (2018). An introduction to text mining: How libraries can support digital scholars. *Qualitative and Quantitative Methods in Libraries*, 7(1), Article 1.
- Weaver, D. A., & Bimber, B. (2008). Finding news stories: A Comparison of searches using LexisNexis and Google News. *Journalism & Mass Communication Quarterly*, 85(3), 515–530. <https://doi.org/10.1177/107769900808500303>

-
- Woolley, J. T. (2000). Using media-based data in studies of politics. *American Journal of Political Science*, 44(1), 156–173. JSTOR. <https://doi.org/10.2307/2669301>
- Youngblood, N. E., Bishop, B. A., & Worthington, D. L. (2013). Database search results can differ from newspaper microfilm. *Newspaper Research Journal*, 34(1), 36–49. <https://doi.org/10.1177/073953291303400104>

Appendix

Codebook with Cohen's kappa

Code	Cohen's kappa
Year(s) news content in the scholarly papers' samples were published	
2010-2021	0.97
2000-2009	0.97
1990-1999	0.95
1980-1989	0.85
1970-1979	1.00
1969 and prior	0.79
Did not disclose dates being studied	n/a
Number of years since the news articles in samples were published and the scholarly papers were published	
0-9 years ago	n/a
10-19 years ago	n/a
20+ years ago	n/a
Did not disclose dates	n/a
Number of news outlets in the scholarly papers' samples	
Set list of news outlets	0.89
1 outlet	1.00
2 outlets	1.00
3-5 outlets	0.96
Used all news outlets in database	0.86
Did not disclose number of news outlets	n/a
Types of news outlets in the scholarly papers' samples	
Legacy newspaper U.S.	0.89
Legacy newspaper non-U.S.	1.00
Legacy news' web U.S.	0.82
Legacy news' web non-U.S.	0.88
Digital native news' web U.S.	1.00
Digital native news' web non-U.S.	n/a
TV news websites U.S.	0.92
TV news websites non-U.S.	0.66
Types of news content in the scholarly papers' samples	
Text-based	0.80
Reader comments	1.00
Photos	1.00
Other	0.92

Resources scholarly papers used to access the samples	
Subscription-based news aggregator databases	
LexisNexis Academic	1.00
Nexis Uni	n/a
Factiva	1.00
ProQuest news database	n/a
Other subscription-based news databases	0.66
Non-subscription-based news aggregator databases	
Clipping service	1.00
Public search engine	0.79
Historical newspaper collection	n/a
Other public databases	1.00
News websites	0.92
Microfilm/microfiche	n/a
Hard copies of newspapers	0.66
Did not disclose resources used	0.83