



Informing Science: the International Journal of an Emerging Transdiscipline

An Official Publication
of the Informing Science Institute
InformingScience.org

Inform.nu

Volume 25, 2022

TRUST IN GOOGLE – A TEXTUAL ANALYSIS OF NEWS ARTICLES ABOUT CYBERBULLYING

Juergen G. Stegmair*	University of North Texas, Denton, USA	jstegmair@twu.edu
Victor R. Prybutok	University of North Texas, G. Brint Ryan College of Business and Toulouse Graduate School, Denton, USA	prybutok@unt.edu

* Corresponding author

ABSTRACT

Aim/Purpose	Cyberbullying (CB) is an ongoing phenomenon that affects youth in negative ways. Using online news articles to provide information to schools can help with the development of comprehensive cyberbullying prevention campaigns, and in restoring faith in news reporting. The inclusion of online news also allows for increased awareness of cybersafety issues for youth.
Background	CB is an inherent problem of information delivery and security. Textual analysis provides input into prevention and training efforts to combat the issue.
Methodology	Text extraction and text analysis methods of term and concept extraction; text link analysis and sentiment analysis are performed on a body of news articles.
Contribution	News articles are determined to be a major source of information for comprehensive cyberbullying prevention campaigns.
Findings	Online news articles are relatively neutral in their sentiment; terms and topic extraction provide fertile ground for information presentation and context.
Recommendations for Practitioners	Practitioners should include online news articles (or systems that deliver information extraction) into the cyberbullying curriculum.
Recommendations for Researchers	Researchers should seek support for research projects that extract timely information from online news articles.
Impact on Society	Online news articles are a valuable source of information about CB and are not as biased or opinionated as often believed.

Accepting Editor Eli Cohen | Received: July 1, 2021 | Revised: November 30, December 22, 2021 | Accepted: December 23, 2021.

Cite as: Stegmair, J. G., & Prybutok, V. R. (2022). Trust in Google – A textual analysis of news articles about cyberbullying. *Informing Science: The International Journal of an Emerging Transdiscipline*, 25, 45-63.
<https://doi.org/10.28945/4894>

(CC BY-NC 4.0) This article is licensed to you under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). When you copy and redistribute this paper in full or in part, you need to provide proper attribution to it to ensure that others can later locate this work (and to ensure that others do not accuse you of plagiarism). You may (and we encourage you to) adapt, remix, transform, and build upon the material for any non-commercial purposes. This license does not permit you to use this material for commercial purposes.

Future Research	Refinement of the terms and topics analytic model, as well as a system development approach for information extraction of online CB news.
Keywords	cyberbullying, educational leadership, text analytics

INTRODUCTION

Cyberbullying (CB), the willful, targeted act of using online communication channels to inflict harm and abuse on individuals, continues as a persistent problem with severe consequences. The phenomenon continues after more than two decades.

In one study, 72% of respondents reported at least one incident of CB in the last year, out of a sample of 1,454 students between 12 and 17 years old (Juvonen & Gross, 2008). CB is a phenomenon of everyday life and needs to be treated as such. Vandebosch and Van Cleemput (2008) created a qualitative study with focus groups under the assumption that ICT and CB are part of the everyday life of youngsters and with the hope of gathering important information about usage and modes of communication. Out of the increased usage of Information and Communication Technology (ICT), violating behavior and increased need for protection and safety regarding CB are a direct outcome.

CB is an interdisciplinary issue, with domain application in public health, education, psychology, computers in human behavior, information seeking, informing, legal issues, criminality, and cyber-safety. Trust in news reporting is eroding, particularly in communication behaviors political such as some of those from former President Trump. Grant (2019, p.1) mentions in an article from Informing Science and Information Technology Education Conference that “the Washington Post Trump Fact Checker has identified more than 10,000 untruths uttered by the president in his first two years in office.” Fake news has become a buzzword in the public mind, further eroding the trust in new media and news organizations. Fake news attribution ranges from manipulating images and videos in a scientific context (López-Cantos, 2019) to combat perceptions about news reporting and social media (Cohen, 2019; Gill, 2019). Online news reporting about cyberbullying increases awareness of cyber-safety issues and online dangers for youth in an increasingly complex online world (Zilka, 2017; 2018).

CB can impact most youth, in primary and in secondary school. Comparing perpetration and victimization in primary school to secondary school, instances of both are seemingly higher in primary school (Dehue et al., 2008). Perpetration was 17.1% in the last class of primary school, versus 13.5% in the first class of secondary school. Victimization was 23.4% versus 18.6%. Occurrences with fatal outcomes are still happening.

One research question worthy of asking is whether there is trust in online news (media) to report objectively about CB and its occurrences. This research addresses the question of whether online news articles can be gathered, transformed, and used to create reliable information to help answer the question of semantic analysis for news articles on CB.

LITERATURE REVIEW

PROLIFERATION OF MEDIA CHANNELS AND TOOLS

Electronic media and online communication channels are ever-changing. Remember ‘MySpace’? Not every online communication channel or social media site stays around forever. Snakenborg et al. (2011) prominently emphasize a letter from a CB victim at the beginning of their article, describing CB occurring on the MySpace account. MySpace is not heavily used in recent CB research articles. Therefore, it will be important to guide CB prevention programs to include various and numerous online communication channels and social media sites. Providing an overview of the past (i.e., MySpace) shows that, while communication channels change over time, behavior and victimization

continue. CB prevention needs to be flexible enough to be evaluated every couple of years to include the newest methods, channels, modes of online communication, and social media behavior. CB prevention specialists should stay current with these ever-changing modes of communication to gain acceptance and credibility with youth. Juvonen and Gross (2008) include in their anonymous web-based survey the following communication channels: message boards (where you can leave messages in an asynchronous fashion to receive answers later), and again, MySpace.com. Both modes of electronic communication are not major outlets in 2020.

ONLINE COMMUNICATION AND SOCIAL MEDIA

Modern communication channels and their use for CB represent one potential factor in the increase in youth suicide rates (Duncan, 2017; Patchin & Hinduja, 2016). The ease of online communication access and availability of social media platforms enable online violations and cyber aggression. Such acts come in many forms: organized actions against persons or organizations in social media for a cause (e.g., #cecilthelion; BBC News, 2015); online harassment and CB against persons of public interest, like actors (e.g., Star Wars Actors; Perez, 2018); or both targeted and general attacks against people, organizations, and groups. For example, Holder (2017) characterizes Trump's use of Twitter as a CB tactic. In this instance, the president's communication includes elements of specification (one person is targeted), demeaning or belittling (nicknames for political opponents are used), an electronic communication channel (Twitter) is used, and there is a power differential (the then president of the United States versus single politicians, members of Congress, governors, or even public media personnel). Such examples have several things in common: the movement away from single communication channels (like Chat) to multiple channels; the ease of use and access to social media platforms; and the occurrence of cyber abuse or aggression with relative ease. While the literature oftentimes explains cyber aggression as a partial outcome of anonymity on the internet (Cohen-Almagor, 2018), it does not fully address cases where the perpetrator is well known and where anonymity is not provided.

Our society has not developed an effective applicable approach to prevent and intervene in cyber aggression behavior. Furthermore, successful coping behaviors and resiliency are tied together and linked to better recognition of cyber aggression behavior.

SOCIAL NETWORK EXPANSION

CB prevention must consider the expansion of social networks. While there was a progression in the past from personal (individual; family-based) to peer-based (school) and then within a community (adult life), such development now includes moving from physical social networking to online social networking (Snakenborg et al., 2011). It is important to acknowledge this development and describe the characteristics of the past vs. present. For example, presenting our best self in the online space (i.e., see 'best life' presentation on Facebook; Picture presentation on Pinterest), distorts how individuals see themselves, and how they are seen by others. Is the online persona 'virtual' in the sense that it does not exist? Can there be conflicts between the actual persona and online representation? These conflicts can be taught and included in CB prevention programs to help children and the young better develop.

It is remarkable that social media expansion also happens in the workplace. By allowing coworkers to be "friends" on, e.g., Facebook, or on a more professional site like LinkedIn, matters of work integrate and intrude in one's social life. This integration has unintended consequences when certain events happen in either work-life or when a certain behavior or situation is observed by colleagues via social media. The social media expansion into the work-life can co-exist or compound a phenomenon of 'cyber incivility', in which individuals at work experience negative email communication, behavior, and mannerisms (Ophoff et al., 2015). Therefore, it is worth understanding the entanglement of those two worlds.

ADOLESCENT ELECTRONIC COMMUNICATION

The electronic communication of adolescents is not open to review or influence by adults (Mason, 2008). Likewise, violations in online communication space are not easily visible to adults. Students and youth prefer not to report CB. Some fear that online aggressors will retaliate, or that parents will get less permissive with internet usage (Snakenborg et al., 2011). Other impressions are that adults, whether teachers at school or parents at home, are either ineffective or uninterested in fighting against the CB case. Parental monitoring reduces the likelihood of youth becoming an online CB aggressor. The question that needs to be asked is the following: what are effective and acceptable ways for parents to monitor and observe online communication? Depending on the age range, even discussion of this topic at home can bring strife into a family. Therefore, the CB prevention program needs to be explicit in how to support parents and students in addressing this topic. For example, parents looking through their children's phones, messages, and browser histories are likely to result in a hostile response.

PRIVACY IN SOCIAL MEDIA

Agosto and Abbas (2017) recognize that to report about CB along with privacy and safety in social media, a counterpoint to popular opinion is needed. Their research shows that teens have noteworthy concerns about online privacy and safety, but not necessarily the knowledge or tools to improve and better the perceived situation. Additionally, considering that the disclosure of personal information online (i.e., who can see your personal pictures, your sentiments), research supports that you make yourself vulnerable by doing so – and you increase your likelihood of becoming a CB victim (Van-debosch & Van Cleemput, 2008).

SAFETY IN SOCIAL MEDIA

The perception of safety in relation to social media is connected to perceived privacy. This aspect of safety manifests itself in several ways. One dimension is that of data safety ('Is my data secured and safe online?'); another is that of personal safety ('Am I safe in conversing and participating in this online channel?', 'Will I encounter personal attacks when posting in this channel?'). Data safety has been under attack in the past; data breaches have occurred, and data have been too easily available for extracting for commercial or other purposes. One such example is the scandal over Facebook and Cambridge Analytical data abuse, which showed how easy it was for a company to gather valuable data for research without the explicit consent of users (Criddle, 2020). Personal safety has deteriorated as online communications often deteriorate into political or personal attacks. This deterioration has caused providers like NextDoor to change policies and reporting guidelines to ensure a civil online communication platform. Other dimensions are emotional and physical safety - you open yourself up emotionally to CB (relationship status, anyone?) and physically to burglars or other online perpetrators (by disclosing information like addresses, modes of occupying your home, vacations, etc.). Improving online safety and privacy of teens is a central tenet of CB prevention programs and requires a whole-system or integrative approach. Lastly, another aspect of feeling safe in an online communication platform is the subjective impression of receiving accurate and true news content, from a platform.

IMPORTANCE AND MEANING OF NEWS

Social media channels are increasingly delivering non-personal news. Prevalent social media platforms like Facebook and Twitter inform people by displaying and distributing news.

Everybody wants to be informed and one source for information is the news. News provides up-to-date information on topics of personal relevancy. Though print media (news, magazines) are still used, consumption happens increasingly online. A study by the Pew Research Center shows that the gap between online and television news consumption is closing; 43% of US adults get their news

online, and two-thirds of Americans use social media as at least a partial news source (Bialik & Matsa, 2017). It is important to have trust in your news source, and the same study mentions that Americans have low trust in information from social media sources.

News can have a devastating impact – examples are catastrophes (9/11, weather, stock market crashes, the message about the death of a loved one), and in macroeconomics, the concept of ‘news shocks’ is well documented (Sims, 2016). News as industry gives “voice to the people” (“vox populi”; Merriam-Webster, 2018), especially on social media platforms. News provides a platform for underrepresented groups and subgroups of people, which are less heard. News reporting can be a change agent in our society (i.e., gender and sexuality identify framing) and can bring help in disaster relief situations (i.e., earthquakes in Turkey, Nepal). News is a balancing force for government and dominant economic forces. The ‘Fifth Estate’, a term used for journalists and bloggers who use non-mainstream media and social media news, serves as a channel to provide inopportune and not readily accepted opinions and news (Fifth estate, 2018). On a micro level, mobile applications like Nextdoor provide a platform for neighborhood-centric information exchange and news about neighborhood happenings.

Reliable and objective news is important especially when media are under attack, such as proclaiming legitimate news media sources an ‘enemy of the American people’ (Grynbaum, 2017) or labeling news media outlets as ‘fake news’ (McCarty, 2018). Alongside critical thinking and fact-checking, news reporting is strengthening the resiliency of the nation and the individual. News reports about cyber aggression will strengthen resiliency for young people, positive examples and reports will lift hope and confidence that cyber aggression can be overcome, and news about training or education initiatives will inform school administrators to potentially mimic actions or inspire local application.

Agosto and Abbas (2017) report that popular media are framing adolescent use of social media as negative with no concern for their online safety. They conducted a mixed-method study in 2017 to describe teens’ attitudes about online safety and privacy. As news reporters are adults, they point out the importance of including youth voices in framing, rather than having these issues defined only presented by adults. As social media use is a CB channel and an important communication outlet for youth, it is of utmost significance to be inclusive.

Smith et al. (2008) point out the importance and influence of media in providing knowledge about CB, framing the issue, and for proliferating items like ‘happy-slapping’ (which then found its way into a popularly used method in cyberbullying), and that this area is not well researched. A more nefarious way to use CB is ‘doxing’, which is a way of disclosing information online to harm individuals or groups to incite murder, as displayed in Afghanistan, where neighbors have disclosed lists of people in their surrounding areas that they perceive as ‘troublesome’ with the effect of a targeted killing wave (Hadid, 2021).

Reports of the UK (Smith et al., 2008), where there is a more vicious and aggressive media culture for online aggression, point to a difference in media framing when looking through a nationality or country lens. A major example of a vicious media landscape is Harry and Meghan Markle’s separation from the UK royal family; ostensibly in part due to the treatment of Meghan in UK print and online media (Dray, 2021).

The inclusion of media in a CB prevention program is warranted and will be enabling for youth who encounter CB reporting in public media. This approach is different from multimedia inclusion in CB prevention programs (Snakenborg et al., 2011). The inclusion of video or webisodes helps make the material more digestible to learners. However, the inclusion of real-life cases and news articles, which provide a reality aspect to students, will be a real impact that such a program can achieve. It is important to include positive news articles, and not to condemn the phenomenon, and to try scaring students. Of course, negative news articles will be helpful to demonstrate the occurrence, scope, impact, and victimization of CB.

Media framing also comes into play when discussing urban myths, legends, and other imaginary stories or events. Agosto and Abbas (2017, p. 357) share the testimonies of students who talk about “creepy old guys” on the Internet or in online video games. One example of an online article is “The Dirtiest Old Men on the Internet” (Boudreaux, 2012). The article mentions examples of offensive and questionable sexual messages to women, which provide suitable education when positioned age-appropriate; however, when a person is characterized as “Like this 49-year-old guy here” (middle of article), the writer of the research article is offended because: (i) he is older than 49; and (ii) unfortunately this type of behavior can be extended from younger males as well. Another passage in the study tells “With the whole creeping thing [accessing one’s personal information without their knowledge], what if some 57-year-old man is creeping on my little sister’s profile?” (p. 360). However, the theme of “Dirty Old Men on the Internet” remains and CB prevention should consider it by informing about it and potentially dispelling the myth.

Media reporting about incidents and how to prevent and cope with CB is important. In a study by Parris et al. (2012), a student reports that it is important to read and hear about CB in the news, especially if it is happening close by. The reporting creates awareness in students and will be helpful to prevent them from engaging in online aggression. Therefore, an important aspect of CB prevention programs is to include news and information about cases in the locality (at their school, in their city/state) to make the issue relevant.

TRUST IN MEDIA

Trust in media and information stems from trusted sources of authority, knowledge, and influence. Wilson (1983) states to the consideration of authority:

People come to have influence over our thoughts in various ways. However, they acquire such influence, as long as we think it right that they should have it, we will be prepared to defend and justify their influence. We will have answers to the question: “Why do you listen to him?” and “Why do you let him influence you so?” (p. 21).

Every adolescent is developing their own bases of authority. This process includes sourcing information from different channels. They derive ways of acquiring news by various examples: news consumption in their home environment (parents, siblings, other relatives); news consumption in their circle of friends and acquaintances (school environment, social environment); and finally, the positioning of media and news acquisition and consumption by teachers and school administration. While school environments are upheld to teach to a critical consumption perspective, the home environment does not necessarily need to be so and can teach and display preferences in media and news consumption, such as political or religious preferences, and educational vs. informational vs. entertaining consumption. Therefore, school environments should provide a critical and informational perspective toward media and news consumption by teaching how to personally evaluate the veracity and authority of news media, and the positioning of items like CB in public media.

The positioning is even more important for social media, as social media is not as much regulated as public news media. While writing this research article, several legislative initiatives are ongoing on both state and federal levels of the US to regulate social media platforms in terms of censorship.

TRUST IN GOOGLE

This study focuses on Google and issues of trust in this search engine. Google is prevalent in everyday life. Especially school children at a young age (primary; secondary school level) encounter Google as a conglomerate of applications (Google Drive, Google Documents, Google Classroom, using Google as a search engine). As much as it is part of their everyday life, young teens do not yet have perceptions of privacy concerns when searching online (i.e., by using a different search engine like DuckDuckGo, which does not store search history), or following public discussions of the omnipotence of social media and tech giants such as Facebook, Twitter, and Google. Especially anti-

trust or economic concerns, which are frequently cited in connection with Facebook and Google, are not seemingly a concern of young teens when they use Google technology and search engines. Therefore, while young teens do not perceivably search for news or information about CB online with the Google search engine, the argument exists that they would perform such a search and information-seeking activity and trust in the returned result set of news.

The viewpoint of older teens (18–19) toward mature people is that older people lack knowledge and grew up before the Internet (or at least before the World Wide Web), therefore they can't change profile settings in Facebook or security settings in other social media or communication channels. Younger children are not educated (knowledgeable) enough and lack forethought, meaning they over-share online and are potentially exploited (Agosto & Abbas, 2017). The viewpoints of younger people toward adults are that much older people are sometimes labeled 'Old Creeps', and that adults, specifically teachers, are unknowledgeable, and can't be trusted.

Adult viewpoints about popular media use by younger persons are that all young people are 'Digital Natives', where 'Digital' implies knowledge and expertise around 'all things' online, and 'Natives' implies the growing-up with online and digital environments and 24-hour communication ability (Agosto & Abbas, 2017). However, 'Natives' is also demeaning and framing the following: they behave like 'natives', i.e., uneducated, uncontrolled, dangerous.

Generational gaps in the usage of and understanding of how and when to use a certain online technology or communication channel are a big block in building bridges and forging trust. Juvonen and Gross (2008) recommend in their results the work particularly required in creating and fostering these linkages, as their participants stated that they must be able to deal with CB incidents themselves, and therefore did not tell a trusted adult. This observation has been brought forward by an anonymous web survey, which did not need parental consent for participation; therefore, it rings particularly true.

RESEARCH QUESTIONS

The following research questions were developed:

- Q1: Is Google a trustworthy search engine that returns objective news articles?
- Q2: What is the role of objective news articles in a comprehensive CB prevention program?

METHODS AND RESULTS

SENTIMENT ANALYSIS AND TEXT LINK ANALYSIS (TLA)

Sentiment analysis, when used in conjunction with text analytics, is a method to determine emotions or sentiments in the body of text. A model is created that identifies the positive and negative words or expressions (phrases). An approach often used is a dictionary-based approach that holds a key list of negative and positive phrases. The identification of negative and positive is a spectrum and can show gradients of sentiments within the spectrum, mostly to show emphasis for positive or negative sentiment. Sentiment analysis is used to show whether customers are reviewing a product favorably; whether a restaurant or movie has been enjoyed; how political positions are being perceived; how employees are satisfied, and so forth. It is usually used in conjunction with other (business) data to support analysis. For the use case here, sentiment analysis is being deployed for news articles to show whether news articles are neutral (objective) in their reporting, or whether they are overly positive or negative about the subject matter. Sentiment analysis is the term being used in the SAS Visual Analytics tool.

Text link analysis (TLA) is a technology that matches patterns in text. This pattern matching is usually engaged to find and extract relationships between concepts in texts. The patterns can be entirely user-defined. Many text analytics tools and platforms come with their own standard (default) pattern

matching libraries or templates for most common use cases. TLA can be supportive of sentiment analysis and is useful in other contexts (for example, to find links and relationships between documents, or for clustering techniques). TLA is used in this research paper to perform sentiment analysis in a different tool (IBM SPSS Modeler). TLA rules can be defined to detect negation or emphasis of concepts. While a dictionary approach of positive and negative terms is an initial approach, it would list terms like ‘satisfied’ and ‘unsatisfied’, but not combination expression patterns like ‘not satisfied’, ‘very satisfied’. TLA rules can express a rule and identify the words ‘not satisfied’ as <negation> + <positiveTerm>. It can also categorize the sentence “This product wasn’t very good” as “topic + negation + positiveTerm”. Therefore, TLA can provide meaning by analyzing the context of terms and the term itself.

These two methods or techniques were used in analyzing texts and news articles in social science contexts. In the past, this analysis was performed manually and involved marking text, cutting out pieces of text, using different coding schemes, or colors. Now, advances in computing and computer-supported text analytics have automated these techniques allowing faster and more efficient application.

Information extraction involves distilling information from text by defining concepts, categories, and terms. Terms are the smallest forms of text analytics (usually a word); at a lower level, text analytics is identifying tokens (which are the parts of language in terms of words, usually surrounded by white space characters like spaces). When analyzing text, it must be broken down into segments for analysis (codes). Upon identifying these codes, a higher categorization can occur to create code categories to summarize further. Topics and themes are created using Natural Language Processing (NLP) methods to identify common strands of information in text; it is another method for extracting information out of text and to provide a link from the text to context outside text. There are two different forms of topic analysis, and they differ in method and application. The first form is topic modeling, which is an unsupervised machine learning technique that involves no prior knowledge of tags or identifiers. The second form is topic extraction, which is a supervised machine learning technique that requires the existence of identifying topic terms *a priori*. The second method is more exact than the first but requires the preexistence of identifiers (MonkeyLearn, 2021).

The entity extraction is one method for creating and identifying patterns of text that can be recognized (extracted) as a pre-defined item (entity). These pre-defined items are usually concepts or objects that readers are well aware of, for example, Person, Place, Company, Address, Date, Phone Number, and Social Security Number. This default list of entities can be extended and tailored to different use cases. The intent is to extract pieces of text and label them (identify them) as concepts so that analytics can be performed. Entities are items of interest upon which analytics can further be enacted, and data interactivity methods (drill down, drill detail, link to outer data) performed.

The categories are a condensation of terms and topics to summarize textual data. In the research paper, the category ‘bullies’ can exist of one-to-many sub-terms. The categories can be expanded into sub-categories and sub-terms.

DATA COLLECTION AND ANALYSIS

An article text body for 92 documents was extracted from a data collection performed for a different project. The sample of 92 documents was random. Data were gathered using Google search queries for online news; the extraction of text for the online news articles was performed manually into an Excel spreadsheet. More sophisticated services and solutions are available for text extraction of online web pages; however, for this purpose, the text body was created as such.

Data were analyzed in two different steps. The first step was to perform sentiment analysis with SAS Visual Analytics to showcase the capabilities and functions of the visual orientation of that tool. SAS

Visual Analytics is of interest as it would provide applicability in a classroom, with a capability of interactively analyzing and researching news articles by students as part of a comprehensive CB prevention program.

The second step was to perform knowledge extraction for concepts and terms with IBM SPSS Text Analytics Modeler. This tool has more extended capabilities to perform category, concept, and text link analysis, and would be more usable to create the output for an online CB research and analytics tool that could be used in the classroom for CB prevention.

The text file with article bodies was uploaded into SAS VIYA© Visual Analytics, and visual text analytics was performed. Even with the standard settings, and using a standard dictionary, the process identifies terms and topics as well as provides the rudimentary first sentiment analysis by expressing the documents into positive, neutral, and negative categories. The sentiment analysis level is based on the content of the document, and document level, rather than sentence or paragraph level. Figure 1 shows the main dashboard of the tool.

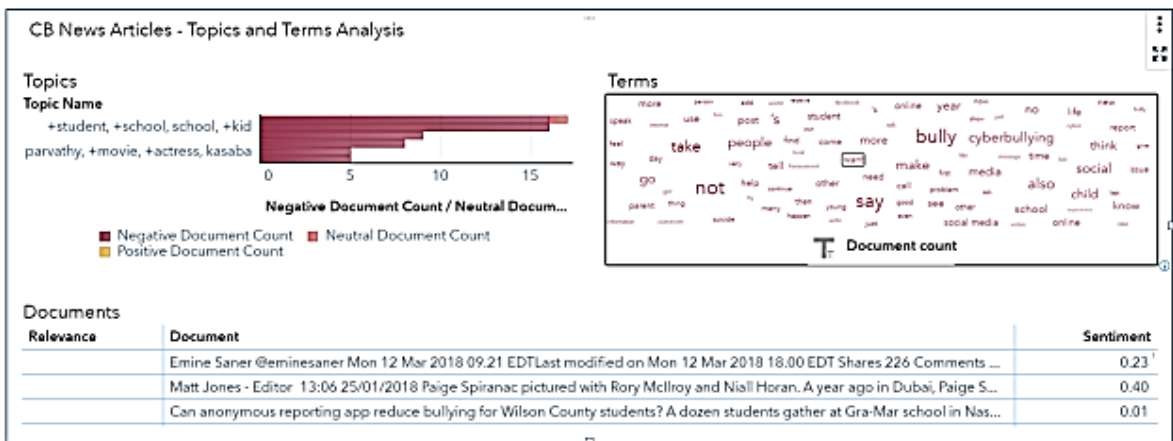


Figure 1. CB news articles - topics and terms analysis - main dashboard information

The terms that were extracted and identified are visualized and expressed as an interactive word cloud in the main dashboard. In addition to visualization for the occurrence of a term, role identification happens for the individual term. Figure 2 shows topics being expressed as a combination of relevant terms.

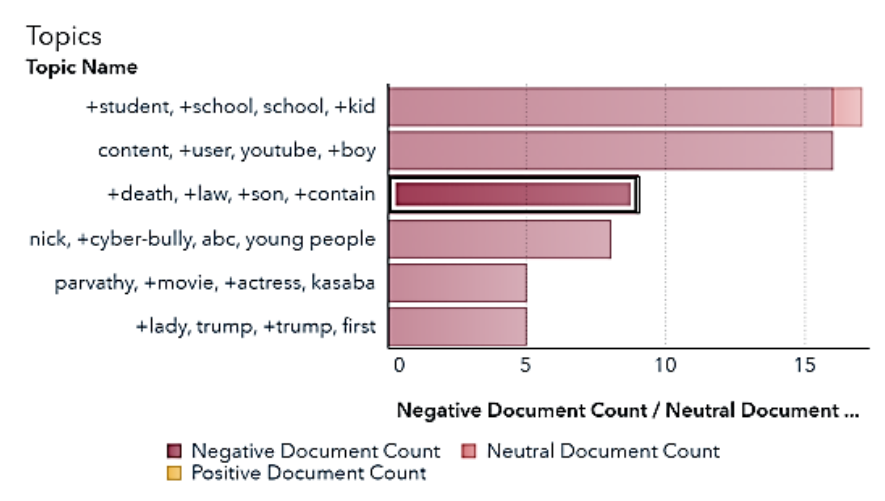


Figure 2. Topic sentiment counts in documents

The topic document frequency table for sentiment analysis (Figure 3) gives an account of negative, positive, and neutral document counts. Additionally, it groups topics to a theme based on co-occurrences and provides a computer-assisted aspect of the definition of themes.

Topic Name	Negative Document Count	Neutral Document Count	Positive Document Count	Document Count
+student, +school, school, +kid	16	1	0	17
content, +user, youtube, +boy	16	0	0	16
+death, +law, +son, +contain	9	0	0	9
nick, +cyber-bully, abc, young people	8	0	0	8
parvathy, +movie, +actress, kasaba	5	0	0	5
+lady, trump, +trump, first	5	0	0	5

Figure 3. Topic document frequency table

Data were collected in the form of URLs, and the text of articles has been extracted manually and stored in an Excel spreadsheet of close to 100 documents. Text mining and analysis were performed using IBM SPSS Modeler Text Analytics 18.2©. The text mining and analysis uses the extraction of terms, concepts, and categories as well as topic link analysis to perform sentiment analysis. Additionally, categorization modeling will be created that allows to categorize news articles and to determine whether a news article is helpful (positive) to be included in a CB information system for youngsters in early secondary school. A total of 92 records (articles) was extracted for this analysis.

CATEGORY ANALYSIS FOR RECOMMENDATION SECTIONS

The category build for the model extracts the categories in Figure 4 for all documents.

Category	Bar /	Selection %	Docs
☑️ 🏠 bullies		85.9	79
☑️ 🏠 family		78.3	72
☑️ 🏠 part		70.7	65
☑️ 🏠 society		64.1	59
☑️ 🏠 people		63.0	58
☑️ 🏠 psychology		63.0	58
☑️ 🏠 academics		57.6	53
☑️ 🏠 human resources		57.6	53
☑️ 🏠 online		57.6	53
☑️ 🏠 school		56.5	52
☑️ 🏠 terrorism subject		55.4	51
☑️ 🏠 geographical location		54.3	50
☑️ 🏠 occupation		54.3	50
☑️ 🏠 crime		50.0	46
☑️ 🏠 consumer electronics		43.5	40
☑️ 🏠 prosecution of law		43.5	40
☑️ 🏠 news		41.3	38
☑️ 🏠 government		37.0	34
☑️ 🏠 suicide		34.8	32
☑️ 🏠 daily		29.3	27
☑️ 🏠 human		29.3	27
☑️ 🏠 movies		28.3	26
☑️ 🏠 january		27.2	25
☑️ 🏠 commercial establishment		26.1	24
☑️ 🏠 sport		26.1	24
☑️ 🏠 state		25.0	23
☑️ 🏠 asia		22.8	21
☑️ 🏠 behavior mechanisms		22.8	21
☑️ 🏠 sex		14.1	13

Figure 4. News articles - initial category build bar - all documents

The top category ('bullies') is not surprising as the news articles are supposed to be about CB. The predefined resources define the 'bullies' categories as shown in Figure 5.

Category	Count	Percentage	Docs
bullies	85.9		79
cyberbullying	78.3		72
anti-bullying	21.7		20

Figure 5. News articles – ‘Bully’ category expansion

There were 6,444 concepts extracted. The top 12 concepts are shown in Figure 6.

Concept	In	Global	Docs	Type
1 children	167	167	41 (45%)	<Unknown>
2 students	162	162	35 (38%)	<Unknown>
3 school	151	151	42 (46%)	<Unknown>
4 people	139	139	50 (54%)	<Unknown>
5 cyberbullying	132	132	63 (68%)	<Unknown>
6 parents	119	119	38 (41%)	<Unknown>
7 kids	81	81	25 (27%)	<Unknown>
8 life	78	78	37 (40%)	<Unknown>
9 social media	76	76	37 (40%)	<Unknown>
10 help	69	69	36 (39%)	<Unknown>
11 girls	65	65	23 (25%)	<Unknown>
12 video	63	63	16 (17%)	<Unknown>

Figure 6. News articles - list of concepts extracted - initially

The expansion of the ‘bully’ concept shows the category structure, which was initially extracted (Figure 7).

bullies	123	79
anti-bullying	18	20
anti-bullying charter	2	2
anti-bullying policies	2	2
bully-free	3	2
bullying act	2	2
bullying behavior	2	3
bullying behaviours	2	3
bullying case	3	3
bullying research	3	3
bullyingrecovery	2	1
cyberbullying	52	72
cyberbullying laws	3	1
face of cyberbullying	2	2
online bullies	4	6
self-bullying	2	1

Figure 7: News articles - expansion of ‘bully’ category - initial

The initial Term-Link Analysis (TLA) (“sentiment analysis”) shows 19 patterns for the combination ‘unknown’:‘conceptual’ and the concept web for it (Figure 8).

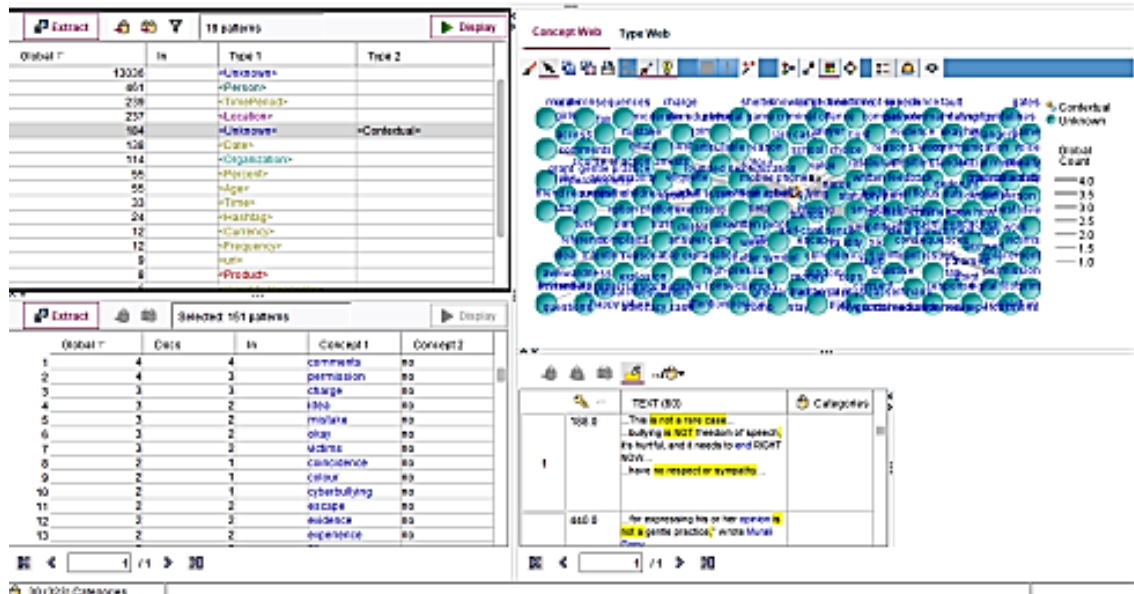


Figure 8. News articles - TLA analysis – initial

SENTIMENT AND TLA ANALYSIS

The sentiment analysis (extraction of the concept ‘no’ for all documents) found **60 documents** (65%) attributed to the concept ‘no’ (**negative**) (Figure 9).

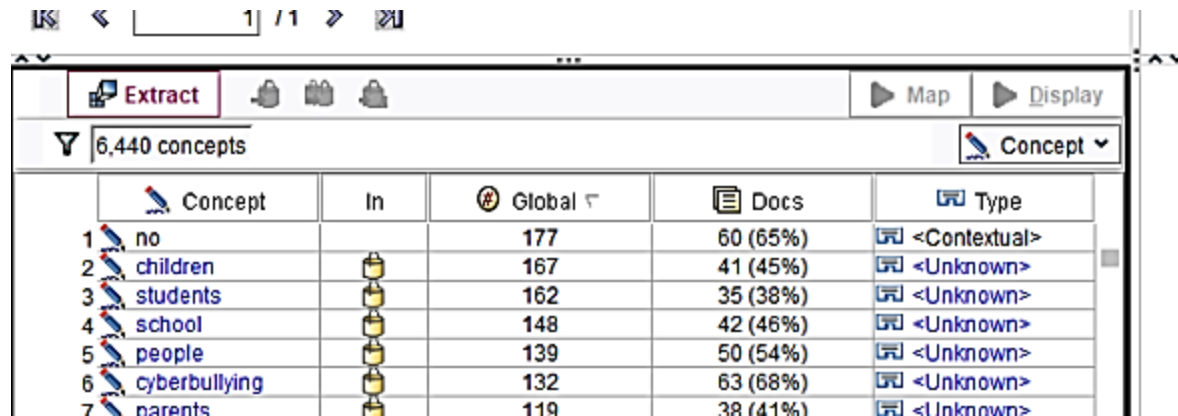


Figure 9. Number of documents contributing to the concept ‘No’

The TLA analysis provides a high count for the type ‘Unknown’ (13,036). An unknown type is a default pre-delivered type for which terms were extracted, but the terms cannot be grouped to a specific type. One challenge to improve the model is to refine terms, concepts, and categories in the interactive workbench to improve the model.

Additionally, only a few item counts (184) show a pattern relationship to the concept ‘no’. No items have shown a relationship to the concept ‘yes’ (both concepts representing here in this case positive or negative sentiments). This finding seems surprising in the first run; however, it makes sense in recognition that the articles in question are news articles and not opinion articles – meaning, news articles are supposed to have greater objectivity (absence of opinion). Therefore, finding a limited number of counts for sentiment analysis seems reasonable.

Figure 10 shows the frequency of extracted types, and type patterns if some have been identified. The current model only identifies 184 occurrences of one type of relation: <unknown>:<Contextual>. Patterns are identified by term patterns (i.e., Phone Numbers; Social Security Numbers; Usernames; Url; Hashtag, etc.).

Global	In	Type 1	Type 2
13036		<Unknown>	
461		<Person>	
239		<TimePeriod>	
237		<Location>	
184		<Unknown>	<Contextual>
138		<Date>	
114		<Organization>	
55		<Percent>	
55		<Age>	
33		<Time>	
24		<Hashtag>	
12		<Currency>	
12		<Frequency>	
9		<url>	
8		<Product>	
6		<Weights-Measures>	
6		<email>	
3		<UserName>	
2		<PhoneNumber>	

Figure 10. Text link analysis pattern table

Out of the TLA pattern type, categories can be created for the defined types. The model included initially the following categories:

- 02_JS_PhoneNumber
- 03_JS_UserName
- 04_JS_Email
- 05_JS_WeightsMeasures
- 06_JS_Product
- 07_JS_Url
- 08_JS_Frequency
- 09_JS_Currency
- 10_JS_Hashtag
- 11_JS_Time
- 12_JS_Age
- 13_JS_Percent
- 14_JS_Organization
- 15_JS_Date
- 16_JS_TimePeriod
- 17_JS_Person

The stream model is the main object for the IBM SPSS Text Analyzer to model the flow graphically (Figure 11).

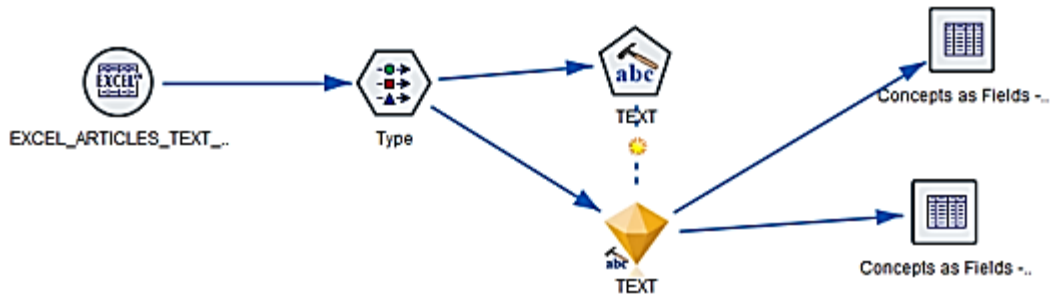


Figure 11. News articles - stream model

HELPFULNESS CATEGORIZATION

To generate a list of articles that is described as ‘helpful’ for funneling into further information consumption for early middle school youth, the list of initial extracted categories was taken and attributed with ‘Helpful’, where ‘Helpful’ is a binary designation; meaning an article is either helpful or unhelpful. This attribution has only been done on the concept level, not on the document level itself.

Of the initial extracted 500 concepts, 336 have been manually attributed (coded) to being ‘Helpful’. The number of 500 is a setting (‘Maximum number of concepts to extract’). In a future step, this number could be increased. However, the created number of concepts was 6,440.

The next step was to create a new category (‘Helpful’) in the model and to assign the designated concepts to that category. The criteria to assign concepts to the new category were relevancy (number of documents > 10) and coded categorization (codebook; manual designation as being ‘helpful’).

After assignment to the new category, and a rerun of the scoring, the extracted model created a new output that makes it easy to select documents (articles) based on the category ‘Helpful’ (Figure 12).

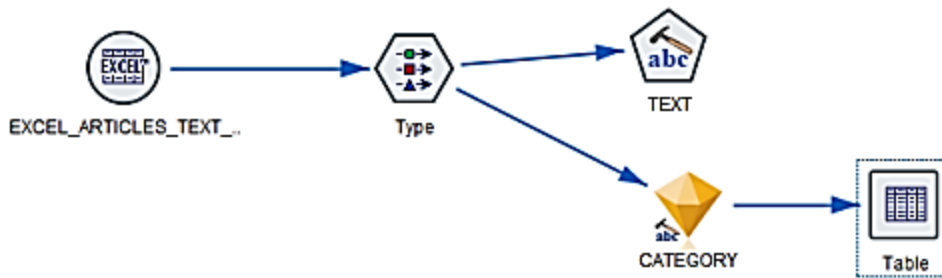


Figure 12. Initial stream model - Category Model

The stream model for categories (‘Category Model’) included the initial extracted concepts and categories only. The second stream model (‘Category Model – HELPFUL ONLY’) was created to create a category specifically for helpful documents only, and an export table based on that. This stream selects only documents that are attributed to ‘helpful’ and then extracts the table only with this one category field. This output table is the desired output set to be used in further information consumption.

The method for creating a new category and attributing documents based on the extracted 500 concepts has limitations (i.e., a single document that reports about ‘Justin Bieber’ might be more interesting to the intended early middle school age group, and therefore also translating into helpfulness for CB prevention and information providing). Future steps would be to present this model to some

early middle schoolers and test run a prototype of the system to identify interesting and helpful documents (concepts and terms).

Furthermore, all extracted concepts can be made available into the extract, which means that concepts such as 'Person' and 'Location' are available for document selection (Figure 13).



Figure 13. Stream model - category model - helpful only

To improve the initial model with a limit of 500 concept extraction, the second run was performed by changing this limit to 1,000. In the first run, 6,440 concepts were extracted in the model, out of which 500 were exported out in the model table export. In a similar fashion, a different model can be created that designates articles determined to be consumed by other groups than students (i.e., teachers, parents, administrators), or differentiation into different age groups for students.

The concepts 'url' and '#' are useful in further analyses as they contain either website URL for helpful organizations or hashtags used in campaigns or organizations. Figure 14 shows the initial extracts for 'url' and '#'.

URI Type /	In	Global	Docs
1 <Age>		55	33 (36%)
2 <Contextual>		177	60 (65%)
3 <Currency>		12	6 (7%)
4 <Date>		138	60 (65%)
5 <email>		6	6 (7%)
6 <Frequency>		12	9 (10%)
7 <Hashtag>		24	11 (12%)
8 <Holiday>		1	1 (1%)
9 <LanguageNegative>		9	2 (2%)
10 <Location>		237	54 (59%)
11 <NewType>		10	8 (9%)
12 <Organization>		114	44 (48%)
13 <Percent>		55	14 (15%)
14 <Person>		451	80 (87%)
15 <PhoneNumber>		2	2 (2%)
16 <Product>		8	6 (7%)
17 <School>		1	1 (1%)
18 <Time>		33	25 (27%)
19 <TimePeriod>		238	67 (73%)
20 <Unknown>		13200	89 (97%)
21 <url>		9	4 (4%)
22 <UserName>		3	3 (3%)
23 <Weights-Measures>		6	5 (5%)

Figure 14. Extract type frequency table

DISCUSSION

While using only the standard delivered elements of the text analytics platforms of SAS Visual Analytics and IBM SPSS Text Miner, sentiment analysis performed with the two tools gained valuable insight into the sentiments of the online news articles. Consistent with the news articles examined, the overall sentiment of the articles was largely objective; as an outcome of this result, online news articles searched and delivered with the Google search engine are categorized as objective. This understanding can help restore trust in online news articles as a neutral source and provides an answer to research question Q1 (“Is Google a trustworthy search engine that returns objective new articles”). Additionally, the information extraction and categorization of online news articles performed using the IBM SPSS Text Miner tool were successful in providing a categorization model for the selected 92 articles that is helpful in critical analysis of the online news articles within a comprehensive cyberbullying prevention program, and this provides an answer to research question Q2 (“What is the role of objective news articles in a comprehensive CB prevention program?”) as it provides objective information to combat this very complex phenomenon.

FUTURE WORK

Extensions of this research work offer opportunities for several future research projects. First, the extraction of a more sizable body of news articles to create a larger text document base is an option, providing the data is readily available. Second, the review of the standard sentiment lexicon used by SAS Visual Analytics will increase the accuracy of the model and will allow sentiment analysis to be fine-tuned toward the domain of CB. Potentially, open-source and existing sentiment lexica could be evaluated and used in the model to further improve accuracy. Finally, a collection of words, phrases, and expressions could be developed that can show young students’ actions and language that are not allowed to be used in civil discourse and should not exist in common communication social media channels and cases that are reported in online news.

SUMMARY

In summary, in the extracted terms and topics in this model (six topics for the abstracts were created), the maximum sentiment scores were between 0.00 and 0.94. One positive document (score: 0.94), one neutral document (score: 0.50), and the rest of the documents (score between 0.00 and 0.49) were classified as being ‘negative’. Obstacles and limitations to this approach include the cost of the tool platform used, as the SAS Enterprise Platform for Visual Analytics is expensive and cost-prohibitive to use for educational purposes in school districts; therefore, low-cost to no-cost alternatives would have to be developed in the future for the visual analysis and interpretation platform. Further analysis and work are needed to identify the classification algorithm and file for the sentiment classification, but the preliminary result of the analysis is showing that online news can be presented in an objective matter to identify – and hand power into the information consumer – sentiment of online news articles. One possible application of this research is the creation of a touch-sensitive dashboard that lets students play and inform about CB.

Additionally, a categorization for the news articles was provided by extracting terms and creating categories. With this method, information can be provided to a CB prevention program and regular teaching activities by bringing CB-related news closer to students and by providing a platform for critical evaluation of many CB prevention-related topics, therefore increasing CB literacy and resilience.

REFERENCES

- Agosto, D., & Abbas, J. (2017). “Don’t be dumb – that’s the rule I try to live by”: A closer look at older teens’ online privacy and safety attitudes. *New Media & Society*, 19(3), 347-365. <https://doi.org/10.1177/1461444815606121>
- BBC News. (2015). *BBC trending – How the Internet descended on the man who killed Cecil the Lion*. <https://www.bbc.com/news/blogs-trending-33694075>
- Bialik, K., & Matsa, E. (2017). Key trends in social and digital news media. *Pew Research Center*. <http://www.pewresearch.org/fact-tank/2017/10/04/key-trends-in-social-and-digital-news-media/>
- Boudreaux, O. (2012). The dirtiest old men on the internet. *buzzfeednews.com*. <https://www.buzzfeednews.com/article/annals/the-dirtiest-old-men-on-the-internet>
- Criddle, C. (2020). Facebook sued over Cambridge Analytica data scandal. *BBC News*. <https://www.bbc.com/news/technology-54722362>
- Cohen, E. (2019). Building an Informing Science model in light of fake news. *Informing Science: The International Journal of an Emerging Transdiscipline*, 22, 95-114. <https://doi.org/10.28945/4486>
- Cohen-Almagor, R. (2018). Social responsibility on the Internet: Addressing the challenge of cyberbullying. *Aggression and Violent Behavior*, 39, 42-52. <https://doi.org/10.1016/j.avb.2018.01.001>
- Dehue, F., Bolman, C., & Völlink, T. (2008). Cyberbullying: Youngsters’ experiences and parental perception. *CyberPsychology & Behavior*, 11(2). <https://doi.org/10.1089/cpb.2007.0008>
- Dray, K. (2021). A comprehensive list of all the s**t Meghan Markle has taken from the British press and public. *Stylist*. <https://www.stylist.co.uk/people/meghan-markle-racist-bullying-tabloids-prince-harry-wardrobe-malfunction-duchess-difficult-examples/342213>
- Duncan, J. (2017). Smartphone, cyberbullying seen as possible causes of rising teen suicide rates. *CBS News*. <https://www.cbsnews.com/news/smartphones-cyberbullying-targeted-as-causes-of-skyrocketing-teen-suicide-rate/>
- Fifth estate (2018). In *Wikipedia*. https://en.wikipedia.org/wiki/Fifth_Estate
- Gill, T. G. (2019). Fake news and Informing Science. *Informing Science: The International Journal of an Emerging Transdiscipline*, 22, 115-136. <https://doi.org/10.28945/4265>
- Grant, A. J. (2019). Ethos, pathos and logos: Rhetorical fixes for an old problem: Fake News. *Proceedings of the 2019 Informing Science and Information Technology Education Conference, Jerusalem, Israel*, 81-91. <https://doi.org/10.28945/4154>
- Grynbaum, M. (2017). Trump calls the news media the ‘Enemy of the American People’. *New York Times*. <https://www.nytimes.com/2017/02/17/business/trump-calls-the-news-media-the-enemy-of-the-people.html>
- Hadid, D. (2021). Amid a wave of targeted killings in Afghanistan, she’s No. 11 on a murder list. *NPR*. <https://www.npr.org/2021/03/21/977797909/amid-a-wave-of-targeted-killings-in-afghanistan-shes-no-11-on-a-murder-list>
- Holder, S. (2017). How the psychology of cyberbullying explains Trump’s tweets. *Politico Magazine*. <https://www.politico.com/magazine/story/2017/07/03/how-the-psychology-of-cyberbullying-explains-trumps-tweets-215333/>
- Juvonen, J., & Gross, E. F. (2008). Extending the school grounds? Bullying experiences in cyberspace. *Journal of School Health*, 78(9). <https://doi.org/10.1111/j.1746-1561.2008.00335.x>
- López-Cantos, F. (2019). The impact on public trust of image manipulation in science. *Informing Science: The International Journal of an Emerging Transdiscipline*, 22, 44-53. <https://doi.org/10.28945/4407>
- Merriam-Webster. (2018). Vox populi. In *Merriam-Webster.com dictionary*. <https://www.merriam-webster.com/dictionary/vox%20populi>

- Mason, K. L. (2008). Cyberbullying: A preliminary assessment for school personnel. *Psychology in the Schools*, 45(4). <https://doi.org/10.1002/pits.20301>
- McCarty, T. (2018). Trump takes war on ‘fake news’ to UK – And tells towering, easily debunked lies. *The Guardian*. <https://www.theguardian.com/us-news/2018/jul/13/trump-fake-news-fox-cnn-theresa-may>
- MonkeyLearn. (2021). Topic analysis: The ultimate guide. *MonkeyLearn*. <https://monkeylearn.com/topic-analysis>
- Ophoff, J., Machaka, T., & Stander, A. (2015). Exploring the impact of cyber incivility in the workplace. *Proceedings of 2015 Informing Science and Information Technology Education Conference, Tampa, Florida*, 493-504. <https://doi.org/10.28945/2248>
- Parris, L., Varjas, K., Meyers, J., & Cutts, H. (2012). High school students’ perceptions of coping with cyberbullying. *Youth & Society*, 44(2), 284-306. <https://doi.org/10.1177/0044118X11398881>
- Patchin, J. W., & Hinduja, S. (2016). Summary of our cyberbullying research (2004–2016). *Cyberbullying Research Center*. <https://cyberbullying.org/summary-of-our-cyberbullying-research>
- Perez, C. (2018). ‘Star Wars’ actress Kelly Marie Tran quits social media after harassment. *New York Post*. <https://pagesix.com/2018/06/05/star-wars-actress-kelly-marie-tran-quits-social-media-after-harassment/>
- Sims, E. (2016). What’s news in news – A cautionary note on using a variance decomposition to assess the quantitative importance of news shocks. *Journal of Economic Dynamics & Control*, 73, 41-60. <https://doi.org/10.1016/j.jedc.2016.09.005>
- Smith, K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: Its nature and impact in secondary school pupils. *The Journal of Child Psychology and Psychiatry*, 49(4), 376-385. <https://doi.org/10.1111/j.1469-7610.2007.01846.x>
- Snakenborg, J., Van Acker, R., & Gable, R. A. (2011). Cyberbullying: Prevention and intervention to protect our children and youth. *Preventing School Failure*, 55(2), 88-95. <https://doi.org/10.1080/1045988X.2011.539454>
- Vandebosch, H., & Van Cleemput, K. (2008). Defining cyberbullying: A qualitative research into the perceptions of youngsters. *CyberPsychology & Behavior*, 11(4). <https://doi.org/10.1089/cpb.2007.0042>
- Wilson, P. (1983). *Second-hand knowledge – An inquiry into cognitive authority*. Greenwood Press.
- Zilka, G. C. (2017). Awareness of e-safety and potential online dangers among children and teenagers. *Journal of Information Technology Education: Research*, 16, 319-338. <https://doi.org/10.28945/3864>
- Zilka, G. C. (2018). E-safety in the use of social networking app by children, adolescents, and young adults. *Interdisciplinary Journal of e-Skills and Lifelong Learning*, 14, 177-190. <https://doi.org/10.28945/4136>

AUTHORS



Dr. Jürgen G. Stegmair is an IT professional, working for the Texas Woman’s University. He received the equivalent of an M.S. of Computing Science from Fachhochschule Würzburg-Schweinfurt-Aschaffenburg, and a Ph.D. in Information Science in 2021 from the University of North Texas. His research interests are information security and education.



Dr. Victor R. Prybutok is a Regents Professor of Decision Sciences in the G. Brint Ryan College of Business, Vice Provost for Graduate Education, and Dean of the Toulouse Graduate School at the University of North Texas. He received, from Drexel University, his B.S. with High Honors in 1974, an M.S. in Bio-Mathematics in 1976, an M.S. in Environmental Health in 1980, and a Ph.D. in Environmental Analysis and Applied Statistics in 1984. Dr. Prybutok is an ASQ certified quality engineer, certified quality auditor, certified manager of quality/organizational excellence, and served as a Texas Quality Award Examiner in 1993. He was also an accredited professional statistician (PSTAT®) by the American Statistical Association. He has authored over 200 journal articles, more than 300 conference presentations/proceedings, and several book chapters in applied and theoretical areas of information systems measurement, quality control, risk assessment, applied statistics, and the instruction of statistics. Awards he received include the 2015 Federation of Business Disciplines Outstanding Educator Award, South West Decision Sciences Institute Outstanding Educator Award, the 2017 American Society for Quality Gryna Award for co-authoring a journal article published in *Quality Management Journal*, the 2018 Decision Sciences Institute Lifetime Distinguished Educator Award, and the 2020 Southwest Decision Sciences Institute Distinguished Service Award.