

## » References to social media based studies using WordStat and QDA Miner

- Al-Rawi, A. (2014). Framing the online women's movements in the Arab world. *Information, Communication & Society*, 17(9), 1147-1161.
- Al-Rawi, A. (2016). News Organizations 2.0: A comparative study of Twitter news. *Journalism Practice*, 11(6), 705-720.
- Al-Rawi, A. (2016). News values on social media: News organizations' Facebook use. *Journalism, Theory, Practice & Criticism*, 17(3), 1-19.
- Al-Rawi, A. (2016). Understanding the Social Media Audiences of Radio Stations. *Journal of Radio & Audio Media*, 23(1), 50-67.
- Al-Rawi, A. (2017). Assessing public sentiments and news preferences on Al Jazeera and Al Arabiya. *International Communication Gazette*, 79(1), 26-44.
- Brennan, R., & Croft, R. (2012). The use of social media in B2B marketing and branding: An exploratory study. *Journal of Customer Behaviour*, 11(2), 101-115.
- Bruns, Axel & Burgess, Jean E. (2011). #Ausvotes: how Twitter covered the 2010 Australian federal election. *Communication, Politics and Culture*, 44(2), 37-56.
- Chen, H. L. (2012). Identifying factors of online news comments. *Proceedings of the Association for Information Science and Technology*, 49(1), 1-4.
- Conway, B. A., Kenski, K., & Wang, D. (2015). The Rise of Twitter in the Political Campaign: Searching for Intermedia Agenda-Setting Effects in the Presidential Primary. *Journal of Computer-Mediated Communication*, 20(4), 363-380.
- Cruz, R. A. B., & Lee, H. J. (2014). The Brand Personality Effect: Communicating Brand Personality on Twitter and its Influence on Online Community Engagement. *Journal of Intelligence and Information Systems*, 20(1), 67-101.
- Chu, K. H., Sidhu, A. K., & Valente, T. W. (2015). Electronic cigarette marketing online: a multi-site, multi-product comparison. *JMIR public health and surveillance*, 1(2).
- Davalos, S., Merchant, A., Rose, G. M., Lessley, B. J., & Teredesai, A. M. (2015). 'The good old days': An examination of nostalgia in Facebook posts. *International Journal of Human - Computer Studies*, 83, 83-93.
- Dexter, S., & Kozbelt, A. (2013, June). Closing the gaps: toward unifying and deepening the study of creativity. In *Proceedings of the 9th ACM Conference on Creativity & Cognition* (pp. 366-369). ACM.
- Groshek, J., & Al-Rawi, A. (2013). Public sentiment and critical framing in social media content during the 2012 US presidential campaign. *Social Science Computer Review*, 31(5), 563-576.
- Groshek, J., & Engelbert, J. (2013). Double differentiation in a cross-national comparison of populist political movements and online media uses in the United States and the Netherlands. *New Media & Society*, 15(2), 183-202.
- Lamy, F. R., Daniulaityte, R., Sheth, A., Nahhas, R. W., Martins, S. S., Boyer, E. W., & Carlson, R. G. (2016). "Those edibles hit hard": Exploration of Twitter data on cannabis edibles in the U.S. *Drug and Alcohol Dependence*, 164, 64-70.
- Luengo, F., Morillo, C., & Yedra, Y. (2017). Categorización de usuarios de Twitter/Categorizing Twitter users. *Revista Tecnológica URU*, (11), 35-44.
- Miller, B., & Behm-Morawitz, E. (2017). Exploring social television, opinion leaders, and Twitter audience reactions to Diane Sawyer's coming out interview with Caitlyn Jenner. *International Journal of Transgenderism*, 18(2), 140-153.
- Nugroho, A., Harwani, Y., Dewita, A., & Sihite, J. (2015). Is It Traditional or Contemporary Marketing Strategy? A Textual Cluster Analysis @MercuBuana\_Reg. *Mediterranean Journal of Social Sciences*, 6(5 55), 26.
- Ruggiero, A., & Vos, M. (2014). Social media monitoring for crisis communication: Process, methods and trends in the scientific literature. *Online Journal of Communication and Media Technologies*, 4(1), 105.
- Settles, P. (2016). *What Goes Up Must Not Come Down: The Tweet Retraction Process of Politicians*. (Unpublished Thesis Project). Western Kentucky University, Bowling Green, KY.
- Sevin, H. E. (2014). Understanding cities through city brands: City branding as a social and semantic network. *Cities*, 38, 47-56.
- Siddiqua, U. A., Ahsan, T., & Chy, A. N. (2016). Combining a rule-based classifier with weakly supervised learning for twitter sentiment analysis. In *Innovations in Science, Engineering and Technology (ICISSET), International Conference on* (pp. 1-4). IEEE.
- Stockemer, D., & Barisione, M. (2017). The 'new'discourse of the Front National under Marine Le Pen: A slight change with a big impact. *European Journal of Communication*, 32(2), 100-115.
- Tse, Y. K., Zhang, M., Doherty, B., Chappell, P., & Garnett, P. (2016). Insight from the horsemeat scandal: Exploring the consumers' opinion of tweets toward Tesco. *Industrial Management & Data Systems*, 116(6), 1178-1200.
- Tucker, I., Goodings, L., Raymond-Barker, B., & Molloy-Vaughan, S. (2015). Social Media and Austerity. *Working Papers of the Communities & Culture Network+*, 5.
- Vasi, I. B., Walker, E. T., Johnson, J. S., & Tan, H. F. (2015). "No Fracking Way!" Documentary Film, Discursive Opportunity, and Local Opposition against Hydraulic Fracturing in the United States, 2010 to 2013. *American Sociological Review*, 80(5), 934-959.

## » References

- Greenwood, S., Perrin, A., Duggan, M., & Pew Research Center (2016). *Social media update 2016: Facebook usage and engagement is on the rise, while adoption of other formats holds steady*.

**DOWNLOAD A TRIAL VERSION AT: [provalisresearch.com/trial](http://provalisresearch.com/trial)**

# ANALYZING SOCIAL MEDIA

## » Analyzing social media data

A parallel is frequently drawn between data and oil, both being valuable resources waiting to be mined. Social media data makes up a large proportion of what is commonly referred to as "big data".

Recent demographics published by the Pew Research Center (2016) indicate that 8 out of 10 Americans using the internet use Facebook and approximately 1 in 4 use Twitter. When you consider that 86% of Americans currently use the internet, you get a better idea of the amount of data out there to be mined. Reddit, the social news aggregator, is the 4<sup>th</sup> most visited site in the US and the 8<sup>th</sup> in the world. This content rich website garnered over 200 million monthly visitors as of Aug 2017.

Social media data is being used widely in research across disciplines, including political science, communications, journalism and business. The impact of social media on the political landscape and the extent of its influence in the campaigning and the election process, as well as the interplay between traditional and social media, are current topics of interest. Social media data can also serve as a barometer for monitoring changing attitudes toward newsworthy or controversial issues.

The results of social media research can be leveraged in business. They can be used as both a means of understanding client needs and developing communication and advertising strategies geared to serving those needs. Social media research can also serve to help determine a company's exposure within the market and be used to track competitors.

Mining social media textual data, like oil, can be made considerably easier if you have the right tools. The text analytics technology of QDA Miner and WordStat gives you the ability to pull and monitor vast quantities of social media data directly from Twitter, public Facebook and reddit pages and RSS feeds, quickly identify keywords and themes, extract topics and categorize data according to pre-established topics. If location data is available, you can even map your results using our unique and easy to use GIS mapping tool.

## » Software features useful for working with social media data

A number of key features in QDA Miner and WordStat are very effective when analyzing social media data.

**Automatic importation and monitoring from Twitter, reddit, RSS feeds and public Facebook pages:** One of the most useful new features in QDA Miner is the ability to import data directly from Twitter, Reddit, RSS feeds and public Facebook pages and monitor your query beyond a single search. Twitter firehouse access, which offers you 100% of the data pertaining to your query, can currently be purchased through third party vendors but can be extremely costly and is often unnecessary for the scope of your project. With QDA Miner you can capture up to 18,000 tweets every 15 minutes at absolutely no data cost.

Your query can be monitored over time using Provalis Research's Web Collector, which allows you to automatically collect data as long as your computer remains on. Set the monitoring parameters of the Web Collector to capture data as often as every minute.

**Automatic variable extraction:** Along with the unstructured text, social media data comes with valuable metadata that can be automatically imported into QDA Miner as variables, which can enable a more comprehensive analysis of your dataset. This metadata varies depending on the source and can include date, Like and Comment counts, Follower and Friend counts, Favorites and Retweets and Upvotes and Downvotes etc.

**Clustering and topic modelling for exploratory text mining:** One of the greatest challenges in working with social media data is its sheer volume. The topic modelling and clustering features of WordStat enable you to quickly see themes emerging from a large dataset. Monitoring relevant themes in the social conversation can provide useful information by revealing dominant or changing opinions on key players, issues, companies or products. It can be useful in brand and crisis monitoring and help inform the trajectory of your research.

**Special character recognition:** WordStat allows you to dictate the type and placement of special characters during text processing. This option allows you to distinguish regular words from hashtags and allows you to measure the frequency of hashtags in your dataset. This information can be leveraged

to measure the popularity of campaigns and to modify hashtags as necessary.

**Crosstabulation with date variables:** Measure the frequency of a topic or a hashtag over time using crosstabulation. This information reveals the ebb and flow of topics and hashtags over time and allows you to pinpoint key dates that may have relevance in your research.

**Geocoding and Mapping:** The metadata associated with your social media data may contain location information. This information may already be in the form of latitude and longitude. If it takes the form of the name of a city, state or country, postal code or even an IP address, QDA Miner's Geocoding feature can help you transform it into latitude and longitude. You can then display the spatial distribution of your codes, topics, keywords and hashtags with our unique and easy to use GIS mapping feature, available in both WordStat and QDA Miner.

**Content analysis dictionaries:** WordStat offers you the possibility to apply a premade categorization dictionary or create a custom categorization dictionary of key words and phrases which allow you to measure specific dimensions of your dataset, including sentiment. Building comprehensive and reusable dictionaries tailored to your subject matter is a great way to automate the categorization process allowing you to save a considerable amount of time.

**Keyword and phrase frequency:** Measuring frequency of keywords and phrases, whether they be product, people or company names, illustrates the online social share-of-voice, valuable exposure information allowing you to gauge the visibility or presence of a person, place or organization amongst its competitors.



## » Examples of studies using WordStat and QDA Miner to analyze social media data

Below are examples of studies that have used QDA Miner and WordStat to help analyze social media data. They illustrate the diversity of domains that are employing social media as a principle data source and the varying methodologies used to analyse this type of data.

Social media is taking on a role of increasing importance in political campaigning and social media users employ these types of platforms to discuss the election issues that matter to them. Bruns and Burgess (2011) monitored the online chatter to identify key themes around the #aussievotes hashtag on Twitter during the 2010 Australian federal election. They used WordStat to extract the most frequent keywords and in turn used these terms to determine five thematic areas. They tracked these themes over time to determine their interconnection with mainstream media coverage and political events.

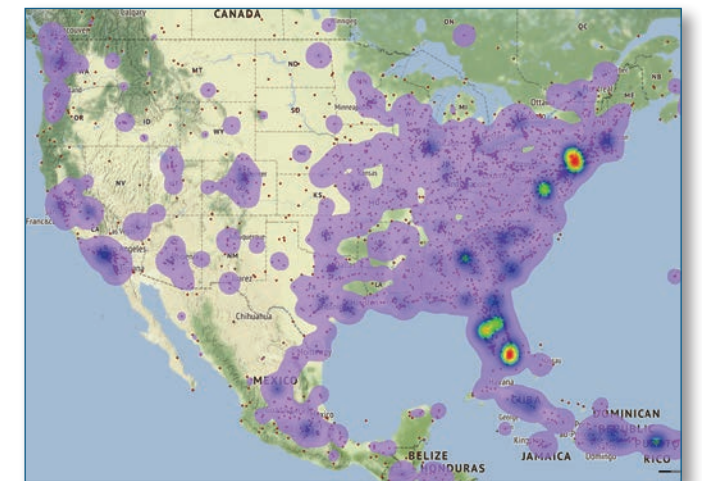
Al-Rawi (2016) in his comparative study of Twitter news used WordStat to identify frequent and cooccurring keywords and topics mentioned in the headlines of over 360,000 tweets on the Twitter pages of 12 English and Arabic news organizations for insight into their news selection, specifically in relation to global proximity of the stories covered.

In their paper which sought to gauge Twitter users' perception of cannabis edibles, Lamy et al. (2016) used QDA Miner to manually code a sample of 3000 tweets for source and sentiment. Intercoder reliability was tested using the Coding Agreement function in QDA Miner. WordStat's frequency analysis was used to identify the most common words attributed to each source and sentiment to determine distinct language patterns with a goal of eventually automating the content analysis process.

Transgender awareness and rights are current issues in both the social and political spheres. Miller and Behm-Morawitz's (2016) conducted a study on audience reception to Caitlyn Jenner's coming out in an interview with

Diane Sawyer through analyzing live-tweet streams produced during the interview. WordStat was used to create a customized theme-based dictionary for classification. The cooccurrence of these themes were then examined in QDA Miner for further meaning and context. They found that despite the negativity expressed in comments on online news stories the live-tweet streams demonstrated an overwhelming positive response to the interview.

Social media, particularly Twitter, is increasingly becoming an indispensable marketing and communications tool. Cruz and Lee (2014) recognize the challenges companies face in developing effective Twitter campaigns. They used WordStat's content analysis capabilities to analyse the Twitter feeds of 23 internationally recognized companies. Terms were categorized based on Aaker's five brand personality dimensions on the brand personality scale. Sentiment analysis was then performed using the Lexicoder Sentiment Dictionary. Results show that word choice and media type are important factors in the success of a campaign and should be taken into consideration by social media managers when developing their marketing and communications plans.



GIS Mapping of 90 000 tweets about the Hurricane Irma.

For more information, go to [provalisresearch.com](http://provalisresearch.com)