

## **THE STUDY OF SENTIMENT ANALYSIS ON THE INFLUENCE OF NEWS HEADLINES ON AUDIENCES**

<sup>1</sup>Tao-Hua Wang, <sup>1</sup>Hao-Chiang Lin, <sup>2</sup>Ting-Ting Wu, <sup>1</sup>Meng-Shain OU, <sup>1</sup>Hong-Da Wang

<sup>1</sup>Department of Information and Technology, National University of Tainan

<sup>2</sup>Graduate School of Technological and Vocational Education,  
National Yulin University of Science and Technology

### **ABSTRACT**

Writing news headlines requires ensuring that not only the essence of a news story is accurately captured but that the headline attracts its audiences. With the rapid circulation of information on the Internet, headlines designed to elicit emotions have replaced conventional headlines to attract “clicks.” In this study, a crawler was used to obtain approximately 4,000 news headlines published within a given month from a popular news website in Taiwan. Word segmentation was subsequently conducted through Jieba, and high-frequency headline words were placed on a list. Ten headlines from five umbrella themes (i.e., headlines news, entertainment news, world news, sports news, and lifestyle supplements) were then selected, and 90 readers from different age groups were invited to label the headlines according to the sentiment evoked. The quantitative results suggested that headlines that elicited happiness and surprise captured the attention of the majority of the participants.

Interviews conducted through the lens of grounded theory suggested that most participants reported overall limited attention to the content of headlines, because they tend to be hyperbolic or misleading. Moreover, the participants prioritized themes of interest over their emotional responses when selecting headlines. They also remarked that emotion- and curiosity-inducing news titles, written primarily to attract clicks, involve commercial intentions because numerous ads are embedded in the news articles, whose visibility depends on their click-through rates.

**Keywords:** Emotion, Grounded Theory, Sentiment Analysis

## INTRODUCTION

The ever-developing Internet technology has saturated contemporary society in information, as well as instigated the digitization of traditional media such as newspapers and magazines. The sheer number of news articles published online, with hundreds of new headlines released every day, ensures that these titles use increasingly sensationalized or interrogative wording to increasing readership. Moreover, the type of sensational wording used in headlines depends on the topic of the story. This study investigated whether sensational online news headlines stimulated the emotions and interest of their audiences, how audiences respond to diverse story topics, and the relationship between online news and commercial activity. Specifically, the following questions were addressed:

1. Do news headlines affect audience emotions? If so, which audience emotions are affected?
2. Do news headlines stimulate the interest of their audience after eliciting their emotions?
3. Do news headlines stimulate audience emotions for commercial purposes?

### *Big data*

Online data, commonly known as “big data” (Bollier, 2010) are increasing rapidly alongside the continual development of the Internet and the growing frequency of mobile devices. Thus, big data are extensively applied; they can be used to

select and analyze the necessary data to identify connections between online information and societal changes, and even detect trends in consumer behavior (Bollier, 2010). Commercial research has shown that big data also provide marketing specialists with information about the behavior of their target customers, upon which subsequent marketing campaigns are based (Boyd & Crawford, 2012).

### *Sentiment analysis*

Emotions originate from the interaction between individuals and their immediate environment, and can be defined in various ways. For example, Lazarus (1991) perceived emotions as psychological processes in which individual changes occur in response to external stimulants. By contrast, Schutz and DeCuir (2002) suggested that emotions change continuously and over time.

Emotions may arise from exciting events or scenarios (Arnold, 1960), body movements (Frijda, 1986), behaviors (Ekman, 1992), or subjective feelings (Russell, 1980). Russell (1980) developed a two-dimensional model of affect that presents eight emotions in four quadrants: pleasure, contentment, excitement, arousal, misery, distress, depression, and sleepiness. Similarly, researchers in Taiwan have typically classified emotions into delight, anger, sorrow, and joy, and Ekman (1978)

organized human emotions into six broad categories: happiness, anger, sadness, disgust, surprise, and fear.

Sentiment analysis is the examination of positive and negative feelings, attitudes, and opinions that are expressed about a given topic (Pang & Lee, 2008; Liu, 2011; Feldman, 2013). A sentiment analysis of textual data reveals the emotional tendencies that are hidden in the text, and generalizes opinions about a topic without using human labor (Liu, 2011; Mullen & Malouf, 2006).

Yeh (2013) argued that textual analysis typically involves three steps: collecting data, processing the data, and analyzing the data and generating results. Data collection can be conducted through a crawler, whereas data processing and analysis are done through text segmentation. Chinese Knowledge Information Processing and Jieba are common word segmentation systems for Chinese researchers, which extract the topic of a text for emotion labeling, and then visually present the data in the form of a sentiment score (Fig. 1).

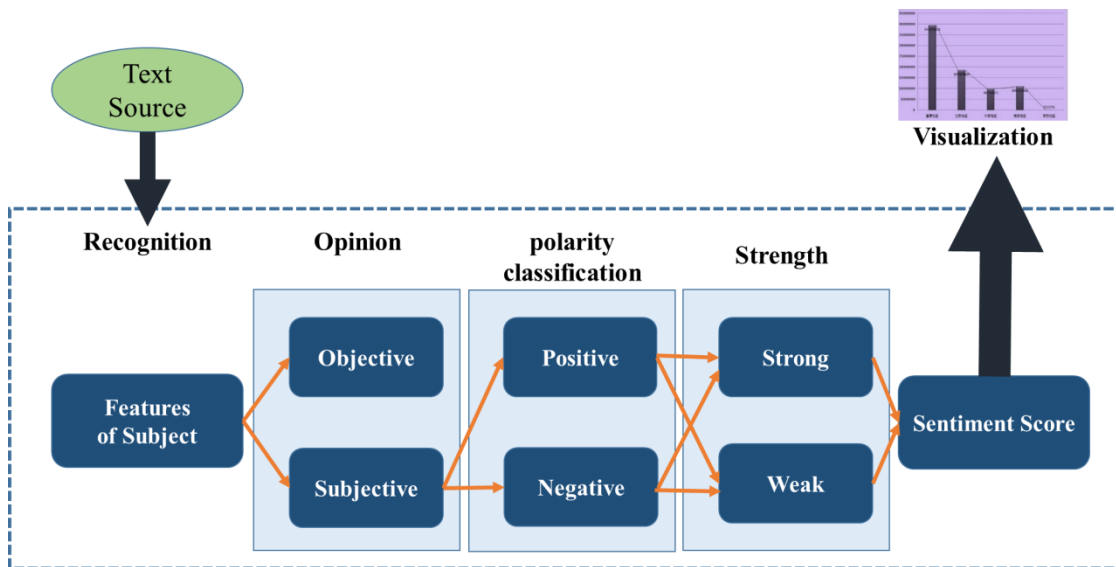


Fig. 1. Sentiment analysis framework (verified from Yeh, 2013)

**Grounded theory**

A qualitative research method, grounded theory states that study problems generally arise from real life situations, and thus should be addressed in real world contexts. After data are organized through comparison, deliberation, analysis,

summarization, and conceptualization, theoretical concepts underlying these data are identified through theoretical sensitivity, which refers to a researcher’s personality and ability to elucidate the subtleties of data meaning (Glaser, 1978). As Strauss and Corbin (1990, 1998) have indicated,

different researchers enter research situations with varying degrees of theoretical sensitivity, depending on their reading and experience relevant to the given area.

One of the analytic stages involved in grounded theory is coding, which dissects and conceptualizes data and subsequently reorganizes the concepts to develop theories about the data (Strauss & Corbin, 1990; 1998). Coding is a linear process that organizes data into structures (Smith & Short, 2001). Additionally, Miles and Huberman (1994) noted that a code is a meaningful unit labeled during the organization of descriptive or inferential data. Generally, coding consists of three steps:

1. Open coding: dissecting, comparing, reviewing, categorizing, and conceptualizing data.
2. Axial coding: analyzing the conditions, contexts, interactions or actions, and results of phenomena of interest, and linking all the categories to recombine raw data.
3. Selective (focused) coding: linking a core category systematically to other categories and verifying the relationship to address categories with incomplete conceptualization.

These three coding processes can be performed interchangeably. Notably, decoding entails conceptualizing and categorizing data, identifying core categories, and establishing theoretical frameworks on the basis of induction and deduction (Pan, 2003).

## **METHODS**

In this study, a crawler was used to capture online news headlines, and word segmentation was subsequently conducted. Term frequencies (TFs) were calculated, and headlines were selected on the basis of high-frequency terms and the opinions of an expert panel. Participants were asked to label the headlines with an emotion according to their subjective perceptions, and the labeling results were then compared. Grounded theory was adopted, and interviews were conducted with three audiences. Coding was later performed on the interview results.

### ***Extracting online news stories***

A crawler was used to extract target online news headlines on the basis of HTML format. HTML comprises such elements as <h>, <body>, and <li>, which are defined as “head,” “body text,” and “unordered table,” respectively. These HTML elements enable the extraction of specific headlines through a crawler.

### Crawler

The crawler used in this study was based on python. After the URLs of target news story webpages were inputted to the program, a webpage extension called InfoLite was used to obtain the HTML elements of the webpages. Finally, scripts were inputted to the crawler to extract headlines from the webpages.

Through InfoLite, the headlines of sample news story webpages were represented by the <article> element, and listed by the <ul> element (Figs. 2 and 3). Next, the URLs of news story webpages were inputted to the crawler, and the <article> elements of interest and their <ul> elements were defined to extract headlines from the webpages.

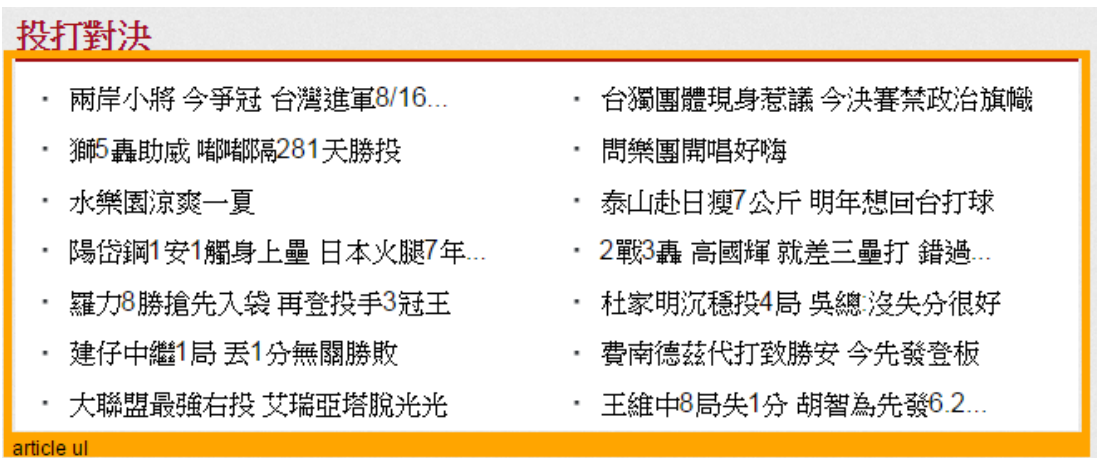


Fig. 2. HTML elements (from <http://udn.com/news/index>)

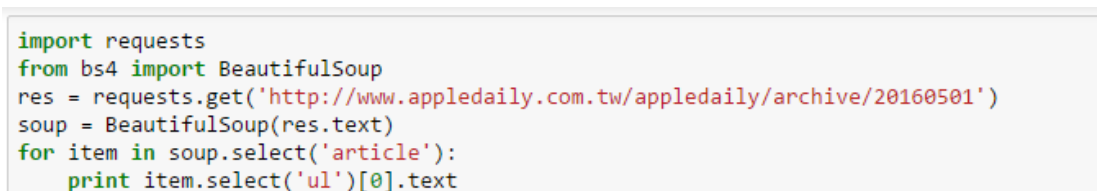


Fig. 3. Extraction of online news stories

### ***Collecting and categorizing news headlines***

In total, 3,876 news articles published online by a Taiwanese media outlet within a given month were collected. Articles published on this site are divided into five themes: headline news, entertainment news, world news, sports news, and lifestyle supplements.

### ***Word segmentation and emotional vocabulary***

Jieba, which is based on open coding, was used to segment the words of news headlines. The TFs of the segmented words were calculated, and the segmented words with emotional content were subsequently selected to identify headlines in which such words appeared. Next, participants were asked to label these headlines with an emotion according to their subjective perceptions.

### ***Emotion labeling***

Ten headlines from each of the thematic categories that contained each high-frequency word were selected. Participants

were asked to label these headlines with one of the six emotions defined by Ekman (i.e., happiness, anger, sadness, disgust, surprise, and fear), or with “neutral” if none of the emotions applied. The data were then quantitatively analyzed and compared.

### ***Grounded theory***

Finally, semi structured interviews were conducted on the basis of headlines labeled with emotional terms. During the interview, participants were asked about the types of emotion that were elicited by the headlines they were presented with, their opinions about the elicited emotions, and whether their reading interest was stimulated by the titles; the interviews results were subsequently coded and conceptualized.

## **RESULTS AND DISCUSSION**

A total of 90 subjects (51 males and 39 females) were recruited to complete a questionnaire on labeling online news headlines with emotional terms. Table 1 presents the five headline themes, and the corresponding emotions that these headlines elicited in the participants.

**Table 1. Online news headline themes and the corresponding emotions elicited.**

Theme	Ranking of the elicited emotion: No. 1	Ranking of the elicited emotion: No. 2	Ranking of the elicited emotion: No. 3	Ranking of the elicited emotion: No. 4	Ranking of the elicited emotion: No. 5	Ranking of the elicited emotion: No. 6	Ranking of the elicited emotion: No. 7
Headline news	Neutral 22.9%	Surprise 19.6%	Anger 15.3%	Sadness 12.4%	Fear 12%	Happiness 9.2%	Disgust 8.5%
Entertainment news	Neutral 52.8%	Surprise 12.7%	Disgust 11.2%	Happiness 10.8%	Anger 6.7%	Sadness 2.6%	Fear 3%
World news	Neutral 31.9%	Surprise 24.8%	Happiness 16.9%	Anger 7.6%	Sadness 6.5%	Fear 6.7%	Disgust 5.4%
Sports news	Neutral 51.4%	Happiness 20%	Surprise 11.9%	Disgust 5.4%	Sadness 4.9%	Anger 3.8%	Fear 2.5%
Lifestyle supplements	Neutral 51.8%	Surprise 18.6%	Happiness 12.8%	Fear 7.6%	Sadness 3.6%	Disgust 3.2%	Anger 2.3%

As Table 1 indicates, the second most frequently elicited emotion after “neutral” was “surprise,” and headlines in all five thematic categories surprised the participants. Surprise arises largely from curiosity; thus, news headlines are typically written to spark the curiosity of their

audiences. Moreover, 58.9% of the participants remarked that online news stories contain emotional elements, and 70% reported an interest in clicking headlines that elicited their emotions.

The participants also commented on the headlines; their comments are summarized as follows:

1. The majority of participants stated that news headlines are characterized by hyperbole, sensationalism, and discrimination.
2. Some headlines are irrelevant to their corresponding articles or lacking in content.
3. Some participants noted that the media purposely uses sensational headlines to pique the curiosity of its audiences, thereby increasing the click-through rates of its articles.
4. Other subjects pointed out that market demand and sales performance pressure

motivate the use of sensational or hyperbolic headlines to capture readers' attention.

***Coding processes in grounded theory***

On the basis of grounded theory, the interview data were analyzed in the following steps: (1) transcribing interview tapes verbatim, (2) simplifying participant statements, (3) conducting coding and establishing core themes, and (4) analyzing coding results. As Table 2 details, the results of the axial coding are the core concepts that were derived by conceptualizing meaningful content segments from open coding.

**Table 2. Axial coding.**

Category	Subcategory	Concept	Meaningful content segment (open code)
Approaches to writing news headlines	News Headline presentation	Exaggerated or misleading news headlines	News articles circulated via social websites tend to exaggerate. <b>A-01-01</b>
			Media reports are filled with hyperbolic claims. <b>B-01-03</b>
			Entertainment news tends to exaggerate and provide information lacking in substance. <b>B-01-12</b>
			Hyperbolic headlines are extensively used in entertainment news articles, but have little relevance to their corresponding articles. <b>B-01-13</b>
			Hyperbolic headlines are extensively used in



			news articles but have little relevance to their corresponding articles. <b>C-01-01</b>
			Despite some hyperbolic news headlines, participants still wanted to click on them. <b>C-01-07</b>
Influence of news headlines on audiences	Feelings about news headlines	Preferences and interests regarding news articles	Preference for news articles that contain certain keywords, or relate to personal interests. <b>A-01-02</b>
	Curious, unusual, or hyperbolic	Arousing audience's curiosity	Preference for curious, unusual, or sad news stories. <b>A-01-04</b>
			Tendency to click on news articles accompanied by curious or unusual headlines. <b>A-01-04</b>
			Most news headlines aroused participants' curiosity but did not elicit their emotions. <b>B-01-10</b>
			Despite some hyperbolic news headlines, participants still wanted to click on them. <b>C-01-07</b>
			News headlines are intended to capture audience's attention. <b>C-01-09</b>
	No specific feeling	No preferred news articles	No specific feeling about news articles, and thus no emotional response to them. <b>B-01-01</b>
			Participants read news articles that concerned them and reported little if any emotional engagement in the articles. <b>B-01-02</b>
			Feel sad when reading some social news stories, but the emotion disappeared afterwards. <b>B-01-05</b>

			Seldom read world news. <b>B-01-08</b>
			Keep up-to-date with news about interesting sports. <b>B-01-09</b>
			Feel little emotion toward entertainment news. <b>C-01-04</b>
			Seldom read world news. <b>C-01-05</b>
			Seldom read lifestyle supplements. <b>C-01-06</b>
New headlines and emotions	News headline (classified by theme)	Feelings toward news headlines	News headlines may elicit readers' emotions. <b>C-01-07</b>
		Feeling certain emotions about news headlines	Social news and entertainment news are more likely to elicit readers' emotions. <b>B-01-04</b>
	Headline news and entertainment news		Entertainment news typically contains hyperboles intended to arouse audience's curiosity. <b>B-01-07</b>
	Entertainment news	Feeling little emotion about news headlines	Feel little emotion toward entertainment news. <b>C-01-04</b>
	Headline news	Eliciting anger	Tendency to feel angry when reading news about murder. <b>B-01-02</b>
			Feel angry or sad about social events but happy about sports events. <b>C-01-02</b>
		Eliciting sadness	Feel sad when reading some social news stories, but the emotion disappeared afterwards. <b>B-01-05</b>
		Eliciting happiness	Feel happy about news stories on the delivery of aid relief in the wake of disasters. <b>B-01-06</b>
News headlines	Commercial	Commercial	News production involves commercial

and commercial activities	tactics used in the writing of news headlines	activities	activities. <b>A-01-06</b>
	Click-through rate and visibility	Advertised via news stories	Commercial products are advertised via news stories. <b>A-01-06</b>
		Click-through rate of product news	Hyperbolic headlining helps to increase the click-through rate of product news. <b>A-01-07</b>
			The media use hyperbolic headlining to increase the click-through rate of its articles. <b>B-01-11</b>
			Ads are typically embedded in webpages and generate revenue when visitors click on the pages. <b>C-01-10</b>
		News articles are essentially commercial	News articles are produced for commercial purposes. <b>C-01-08</b>

Table 3 outlines the results of selective coding, which generated the core categories obtained by conceptualizing the hierarchical relationships between the categories from axial coding. These core categories represent the themes that emerged from the interview responses.

**Table 3. Selective coding.**

Core category	Main category	Cluster
News headlines	Approaches to writing news headlines	News headline presentation
	Influence of news headlines on audiences	Feelings about news headlines
		Curious, unusual, or hyperbolic
		No specific feeling
News headlines and emotions	Thematic classification of news headlines	
Commercial tactics	News headlines and commercial tactics	Commercial tactics used in the writing of news headlines
		Click-through rate and visibility

In short, interview responses were summarized into two core categories: news headlines and commercial tactics. News headlines comprises the main categories of “approaches to writing news headlines” (which focuses on news headline presentation), “the influence of news headlines on audiences” (which focuses on feelings about news headlines), and “news headlines and emotions” (which focuses on the thematic classification of news headlines). By contrast, commercial tactics comprises the main category of “news headlines and commercial tactics,” (which focuses on the clusters of commercial tactics used in the writing of news headlines) and “click-through rate and visibility.”

***Analysis of the axial coding results***

Three conclusions were drawn from the axial coding results: (1) the way news headlines are written affects audience emotions; (2) news headlines may trigger emotional responses from audience; and (3) the way news headlines are written involves commercial interests.

- (1) The way news headlines are written affects audience feelings

The presentation of news headlines determines an article’s first impression, and effectively enticing audiences is the key to writing successful headlines. A news title, in addition to providing the most important point in a report, often intends to attract

readership through the use of hyperbole, sensationalism, and suspense, among other literary devices. However, if such writing skills are applied injudiciously, headlines may seem misleading or exaggerated. Some participants in this study reported that they located news stories by theme or keyword instead, because headlines tend to exaggerate.

(2) News headlines may trigger emotional responses from audiences

News reports vary by content and theme. Thus, their headlines are typically written to elicit readers' emotions in the shortest time possible and entice them to read the stories. Some participants reported that they felt angry or sad about headline news on events, such as murders, but felt happy reading stories from the entertainment news, world news, sports news, and lifestyle supplements categories. Others remarked that their interests, as well as the content of the news stories they read, determined their emotional responses such as happiness, sadness, and surprise.

(3) The way news headlines are written involves commercial interests

When an event emerges, it is frequently reported by multiple media outlets. These sometimes diverse sources provide carefully crafted headlines to vie for readership on the event, which prompts advertisers to pay for ads on the webpages of the articles. When a user clicks on the titles of the articles, numerous ads appear (e.g., in the form of

pop-up pages or in-text images) to promote products or services, thereby creating financial benefits for both advertisers and the host websites.

## CONCLUSION

The three research questions proposed in this study were addressed according to the quantitative and qualitative results:

(1) Do news headlines affect audience emotions and what audience emotions are affected?

The questionnaire results suggested that news headlines affected participants' emotions. Primarily, participants reported neutral emotional responses to headlines of all topics, although feelings of surprise, anger, and fear were also frequently elicited by the news titles of headline news. Additionally, happiness, disgust, and surprise were elicited by a smaller share of headlines in entertainment news; happiness, surprise, and anger were elicited by world news headlines; and happiness and surprise were elicited by headlines in sports news and lifestyle supplements. Furthermore, most participants indicated a strong preference for happiness and surprise-inducing headlines. Finally, the interview results revealed that the themes and content of news stories determined participants' emotional responses.

(2) Do news headlines stimulate the interest of their audience after eliciting their emotions?

According to the questionnaire results, approximately 60% of the participants indicated that online news headlines contain emotional elements, and 70% reported that they tended to click on headlines that elicited their emotions. The interview results similarly revealed that participants preferred emotion- and curiosity-inducing headlines, although they found them irrelevant later.

(3) Do news headlines stimulate audience emotion for commercial purpose?

Some participants provided comments in the questionnaire, stating that news headlines are hyperbolic or sensational because of market demand and performance pressure. Again, the interview results echoed the quantitative data, suggesting that because online news articles typically have a large readership, ads can be placed in the articles for product marketing. Overall, the participants argued that headlines are written in a manner that attracts clicks, thereby delivering the embedded ads, generating revenue for the media that own the articles, and promoting a variety of products.

## REFERENCES

Pan HL. Approaches of Educational Research: Concepts and Applications. Taipei: Education Books; 2003.

Yeh, N. C. (2013). Investigating consumption trends: A sentiment analysis of big data in social media. Retrieved from: <http://ieknet.iek.org.tw/BookView.do?rptidno=129080477>

Arnold MB. Emotion and personality. New York: Columbia University Press; 1960

Bollier D, Firestone CM. The promise and peril of big data. Washington, DC: Aspen Institute, Communications and Society Program. 2010.

Boyd D, Crawford K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*. 2012; 15(5):662-679.

Ekman P. An argument for basic emotions. *Cognition & emotion*. 1992; 6:3-4:169-200.

Ekman P, Friesen W. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press; 1978.

Ekman P. Facial expression and emotion. *American psychologist*. 1993; 48(4): 384.

- Feldman R. Techniques and applications for sentiment analysis: Communications of the ACM, 2013; 56(4), 82-9.
- Frijda NH. The emotions: Studies in emotion and social interaction. Paris: Maison de Sciences de l'Homme; 1986.
- Glaser B. Theoretical sensitivity. Mill Valley, CA: Sociology Press; 1978.
- Lazarus RS. Cognition and motivation in emotion. American psychologist, 1991; 46(4): 352.
- Liu B. Opinion mining and sentiment analysis. Web Data Mining. Springer Berlin Heidelberg. 2011; 459-26.
- Mullen T, Malouf R. A Preliminary Investigation into Sentiment Analysis of Informal Political Discourse. AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs. 2006; 159-62.
- Pang B, Lee L. Opinion mining and sentiment analysis. Foundations and trends in information retrieval. 2008; 2(1-2):1-135.
- Russell JA. A circumflex model of affect. Journal of personality and social psychology. 1980; 39(6):1161.
- Russom P. Big data analytics. TDWI Best Practices Report. 2011; 4:1-35.
- Scherer KR. Vocal affect expression: a review and a model for future research. Psychological bulletin. 1986; 99(2):143.
- Schutz PA, DeCuir JT. Inquiry on emotions in education. Educational Psychologist. 2002; 37(2):125-34.
- Strauss A, Corbin J. Basics of qualitative research: Grounded theory procedures and techniques. Newbury Park, CA: Sage; 1990.
- Strauss A, Corbin J. Basics of qualitative research: Techniques and procedures for developing grounded theory (2nd ed.). Thousand Oaks, CA: Sage; 1998. Miles M, Huberman M. Qualitative data analysis: An expanded sourcebook (2nd ed.). Thousand Oaks, CA: Sage; 1994.
- Smith, C, Short PM. Integrating technology to improve the efficiency of qualitative data analysis: A note on methods. Qualitative Sociology. 2001; 24(3):401-7. Retrieved January 7, 2005, from Academic Search Premier database.