

Linguistics and Literature Review (LLR)

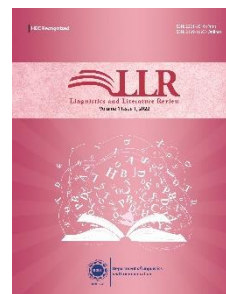
Volume 9 Issue 2, Fall 2023


ISSN(P): 2221-6510 ISSN(E): 2409-109X

Homepage: <http://journals.umt.edu.pk/llr/Home.aspx>



Article QR



- Title:** Analyzing Recent (2019) Kashmir Socio-Political Issue: A Voyant Sentiment Analysis of Tweets
- Author (s):** Saba Zaidi¹, Shaista Allahdad²
- Affiliation (s):** ¹Department of English, Foundation University, Rawalpindi, Pakistan
²Baluchistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan
- DOI:** <https://doi.org/10.32350/llr.92.02>
- History:** Received: February 21, 2023, Revised: June 21, 2023, Accepted: August 15, 2023, Published: September 30, 2023
- Citation:** Zaidi, S., & Allahdad, S. (2023). Analyzing recent (2019) Kashmir socio-political issue: A voyant sentiment analysis of tweets. *Linguistics and Literature Review*, 9(2), 20–46. <https://doi.org/10.32350/llr.92.02>
- Copyright:** © The Authors
- Licensing:**  This article is open access and is distributed under the terms of [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)
- Conflict of Interest:** Author(s) declared no conflict of interest



A publication of

Department of Linguistics and Communications Institute of Liberal Arts
University of Management and Technology Lahore, Pakistan

Analyzing Recent (2019) Kashmir Socio-Political Issue: A Voyant Sentiment Analysis of Tweets

Saba Zaidi^{1*} and Shaista Allahdad²

¹Department of English Foundation University, Rawalpindi, Pakistan

²Baluchistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan

Abstract

Computational Textual Analysis (CTA) is an effective way to analyze large texts by incorporating computational tools, such as, Voyant. The current study takes computer-assisted textual analysis (mixed method) to investigate the sentiments on tweets through Voyant computational method, in order to grasp the emotions of the global public who have tweeted about the recent (2019) Kashmir issue. This research is based on the conceptual framework of Ortony, Clore, and Collins (OCC) model and the keyword approach. For this purpose, different tweets have been analyzed to check the level of sentiments as either positive, negative or neutral. The findings suggested that sentiments were neutral towards the issue of Kashmir and negative towards the Indian government. Voyant has also presented the word count, density, and correlation of phrases within the larger text context. Voyant procedures like; cirrus, trend and summary showed the results based on quantitative and qualitative measures. These toolsets are easy to use without any programming skills and seem to be the best for researchers of social sciences and humanities who are trying to work in digital humanities. The study also recommends Voyant as an operative tool for textual analysis for Computational Linguists, Postmodernism, Critical Discourse Analysis, English Literature, History, Sociology, Theology, and any other field of knowledge that falls under the domain of digital humanities.

Keywords: sentiments, computational textual analysis (CTA), digital humanities, Voyant, Twitter

Introduction

The spread of computer-mediated communication in today's arena is at its peak. People consider it their social responsibility to share every remarkable to tedious detail on social platforms via their smart screens

* Corresponding Author: sabaazaidi@gmail.com

(Mischaud, [2007](#)). They spend their time in argument and virtual world delivery by home entertainment while literature is just a fraction of the available spectrum. In the current era of digital culture where everything is born-digital, even the life of an average individual has potentially transformed. Today's literature is shaped into electronic literature. The digital-born literature celebrates itself as a postmodern art with contemporary roots in technology and social interaction. Hence, the recent generation have entered the age of digital modernism. Digital modernism refers to the strategy of adapting literary modernism into new media literature and sharing diverse groups of digital work (Pisarski, [2017](#)). It works in a broader spectrum; the background of digitalization in humanities is important to understand this study's main purpose. This study deals with digital humanities and how it has transformed literature as an electronic entity with a blend of computational tools to better understand digitalization or digital modernism.

Digital Humanities

Digital humanities, also known as computing humanities, takes computer-generated text as a literary piece. It is a field of study, research, teaching, and invention that is concerned with the intersection of computing and disciplines of the humanities. It is methodological by nature that, involves analysis, synthesis, and presentation of information in electronic form. Digital humanities work through digital platforms such as, hypertext, hyperlinked, via the web, across various disciplines to contribute to our computing knowledge. These platforms are social sites, namely blogs, Facebook, Instagram, Twitter, Yahoo, and Google. The generated text is a digital corpus (Gentzkow et al., [2019](#)). Briefly, the digital spectrum has become an integral part of our existence. Moreover, the justifications down below shed light on how the realm of literature is both created and dissected within the framework of digitalization and technology, which are intertwined in this process.

Technology

Technology has manipulated its way into our lives. It has contributed to social shaping and social construction of our perceptions (Mackay & Gillespie, [1992](#)). Technology has many users and likewise, a large number of readers that have determined its function. The apprising of this function

is instrumental towards the shaping of any society and vice versa (Mackenzie & Wajchman, [1999](#)).

Social Sites

A turning point in the history of telecommunications (Mackenzie & Wajchman, [1999](#)) emerged with the advent of mobile phones and smart screens. Cellphones are an essential part of today's life. They have the most necessary medium of communication for adolescents. It has made the accessibility of the internet affordable and easy. Almost every mobile has social apps, namely Facebook, Twitter, Yahoo, Gmail, Instagram, and WhatsApp with daily visitors on these sites. These apps allow their users to create a public profile and interact with other users on the website. Mobile technology devices allow users to have relationships with information in their own way. This makes the learning experience more fruitful by relating new information to the old information that is already known.

Twitter

The emergence of social media has given web users a platform for expressing and sharing their thoughts (Saif et al., [2012](#)). Among many social apps, Twitter is playing a prominent role in our day-to-day social life. Twitter is a massive social networking site turned towards fast communication. Twitter has nearly 600 million users with over 250 million tweets and with 400 million 140 characters every day. It has been voted in March 2007 as the "Best of the webs" (Mischaud, [2007](#)). Twitter gained popularity in 2005-2010. It enabled the service in the year 2006 with an option, 'status update', and fusing instant messaging. Twitter speaks for people in a person's network that are known and unknown (Pontin, 2007). It works in 140 characters. Twitter with an extensive audience has wider ways of use. Twitter is wildly by most new writers (accomplished or amateurs) who choose a digital platform for their literary piece to be produced for connectivity, subjectivity, and self-reflexivity.

Microblogging

Twitter contains a very large number of short messages created by the user on microblogging platforms. Microblogging platforms like Twitter are used by different people to express their opinions regarding different topics (Pak & Paroubek, [2010](#)), such opinions contain different kind of sentiments about the topics. The sentiments could be positive, negative or neutral;

therefore, Twitter as a micro-blogging website provides a rich source of data for opinion mining and sentimental analysis.

Sentiment Analysis

Every tweet has some sentiments regarding the topic, event, trend, brand or object. Sentiment analysis (SA) is a process of identifying and differentiating users' feelings via their opinions. It is a text classification, which extricates text-based sentiment with the help of different levels of analysis and provides different, methods for these levels. It has its own programming language; Natural Language Processing (NLP). NLP is a computational software which deals with human-computer language interaction (Devika et al., [2016](#)). The classification of Twitter sentiments of thousands of people is not easy without the help of supervised computer programs.

It is challenging to produce something whole in 140 characters. The typical length and irregular structure of a tweet make it demanding. Thus, the form of the Twitter is rigid to 140 characters. Twitter takes different social cultures and literary forms with a mixed composition of genres. It is important to keep the subjective aspect in digital modernism yet maintaining the essence of creativity within 140 characters has limited the literary writers. Tweet to retweet is a way to express anything within the subjective spectrum. Collectively, these tweets and retweets become a large corpus text and demands a critical analysis. The critical analysis of a large number of tweets is essential because it gives a broader context of the text. It also gives voice to the sentiments of the larger audience. This kind of critical analysis helps to examine the cultural and ideological perceptions of people around the globe.

Textual Analysis

The tweets are digital texts (corpus text), and this kind of produced text, in any spectrum goes under the process of textual analysis (TA). Textual analysis tries to understand the meaning of the text, influential variables outside the text and critical evaluation of the context of that text. The textual analysis involves the interpretation of text or multiple texts to comprehend the cultural influence on the text, the denotative and connotative meaning presented in the particular text, and the historical and political agencies supporting the concerned text. Textual analysis is found beneficial in both qualitative and quantitative data analysis. The overall purpose of the textual

analysis is to communicate the present problems of the society and provide a constructive debate on these concerning issues.

Computer-Assisted Text Analysis

Textual analysis has advanced its ways with the emergence of computational science and different computer programs and application software. Computer programs help in examining thousand to millions of text at once. This has given birth to the genre of “-distant reading-”, which is opposite to the traditional method of reading, also known as “-close reading-.” Computer scientists take human judgment more seriously and design their software models with the “-golden standards-” of human intelligence. Digital texts like tweets are a set of algorithms for which manual analysis is difficult and time consuming. Computational text analysis can analyze large data of digital discourse faster and even better than any human doing it manually. The supervised models of the software show less biases on behalf of the researcher or investigator. Computational textual analysis (CTA) is a core component of digital studies, mostly in digital humanities.

Key Factors for Computational Text Analysis

- The text project must be digital with the curation of data base.
- The data must be composed and organized in natural language.
- The research question/s must be clear for digital text analysis.

The digital text undergoes several phases in computational text analysis including text collection, text summary, cleaning, phasing and text visualization. Thus, for conducting analysis and visualization, the computational tools must be rightly available for the usage.

Voyant Tools

The essence of digital humanities (DH) or computing humanities (CH) is that it is convivial to different sciences and multiple interpretations from all the literary forms. The use of supervised software models has proved to be beneficial across various disciplines. It has given greater relief to the social sciences domains as well. One of the friendly use models is Voyant, which is used for computational textual analysis. It is a suggested software for social science scholars who intend to do any kind of textual analysis. Voyant textual analysis can identify human emotions used in the text from

easy using tools. Thus, the current study intends to fill the gap of computational textual analysis with the help of tools like Voyant to analyze Twitter based sentiments, for better comprehension of human sentiments in a broader context.

The Scope of Present Study

The current study intends to take a number of tweets around the world with 4357 words on the current Kashmir socio-political crisis (2019) which has recently employed Twitter with thousands of tweets. People around the world have displayed their sentiments for the brutality that has undergone in Kashmir. Socio-political issues are the reality of our existence and literature is the reflection of that reality. The reality of the surroundings takes the urge for intellectual interference, such kind of interference aims for reasoning, interpretation and evaluation. Thus, this study provides the analysis for the evaluation of a bigger corpus text that manually would have been time-consuming and exhausting. For the sentiment analysis the study limits itself to sentence level and N-gram sentiment analysis. The timeframe dedicated to this research is from February 2019 - May 2019.

Sentence Level of Analysis

The sentence level checks whether each sentence expresses a positive, negative, or neutral opinion. This level is the subjective classification of the present subjective view point of the objective sentences.

N-gram Sentimental Analysis

N- Stands for words or tokens N-grams are the sequence of words in a text or speech. These words can be letters, syllables, and phonemes. It makes use of four types of lexicon, namely sentiment phase, sentiment strength, aspect, and expectation lexicon. This study looked for sentiment phrase lexicons and sentiment strength. Furthermore, this study employed computer-assisted textual analysis as a research methodology to conduct the analysis of the selected corpora.

Research Objectives

Objectives of this research are stated as follows:

- To examine and explore Voyant tools as an effective starting point for computer-assisted text analysis in the field of humanities and social sciences.

- To evaluate the supportiveness of computational textual analysis to grasp and understand the human sentiments.
- To analyze the current Kashmir socio-political calamity under the pining of sentence level and N-gram sentiment analysis.

Research Questions

This research tends to answer the following questions:

- How effective it is to use computer-assisted software for textual analysis?
- How much Voyant is supportive of computational textual analysis in order to grasp human sentiments in the text?
- What is the output of the current Kashmir socio-political calamity from the perspective of sentence level and N-gram sentiment analysis?

Rationale

This research takes the Voyant tools for digital corpus and intends for computational textual analysis in order to analyze the sentiments. Textual analysis is a basic component of the qualitative research in the domain of social sciences and humanities. Previous researchers of literature and linguistics have used different methodologies for the textual analysis. However, the manual process for both approaches was time consuming and difficult until the emergence of computational textual analysis. The only part about the computer-assisted work is that it required expert IT skills or guidance, especially for the scholars of social sciences who were interested to take steps in the field of digital humanities. This issue was resolved with easy programming software like Voyant, which gave a broader analysis of the apparent 5 skins in just a jest of time. Thereby, this study has focused on how the scholars and researchers of humanities could use computer-assisted supervised models, which may give better results, clear from the biases of researchers.

Significance

The current research deploys an exploratory approach, hence, the findings have added valuable results to the scope of digital humanities in literature and linguistics. It has also broader the dimensions for digital modernism as it may encourage researchers who are interested in using easy programming models for textual analysis. The computer has become an

essential need for almost every discipline for diverse purposes; thus, the findings of this research are beneficial for all disciplines as a useful investigation resource.

Delimitation

This study delimits itself to the textual analysis and beginner level of -Voyant tools. Purposely, this study has only looked for sentiments in the corpus. The study is also delimited to the five beginner tools of Voyant for the analysis of the current (2019) Kashmir socio-political crisis. However, further research can carry forward the study by using other advanced Voyant tools with another corpus.

Literature Review

Computational textual analysis is taking its leap in the study of literature and linguistics as an authentic approach and so far the outcome has shown supportive results. There are three main classification levels in sentiment analysis, documental level, sentence level, and aspect level. All of the levels determine the sentiments to be either positive, negative or neutral. The sentence level aims to classify the emotions of each sentence. Wilson et al. (2005) in their study directed that document and sentence levels are the same and that sentences are short documents. Thus, the sentiment expression doesn't have to be subjective in nature, sentence level can extract sentiments from objective fractions and classify it in subjective presentation. As far as the method for the sentiments are concerned, Go et al. (2009) explored, the difference of N-gram with other supervised methods like naïve bayes, maximum entropy and support Vector machine, which, concluded that N-gram improves the sentiment classification accuracy of tweets. Similarly, Barbosa and Feng (2010) argued, though N-grams on tweets can hinder the classification performance because of the large text and microblogging features like retweets, hashtags, and replies. Microblogging features like hashtags, and abbreviations emoticons have enhanced the sentiments of the tweet. Pang and Lee (2004), in a likewise manner conducted a study and, explained that why social scientists should adapt powerful tools that are provided by social scientists to cater to social science-related problem and analyze bug texts with the help of computer-assisted textual analysis.

Evans (2009) in her study, analyzed the demarcation between science and non-science public discourse with topic modeling techniques on

thousands of documents, which were published in America from 1980-2012. The supervised computer model analysis matched the documents, historical and geographical distribution, and analyzed the increase and decrease demarcation. The computational textual analysis with N-gram showed a dramatic increase in the demarcation of language between 1980 - 2012.

Supervised machines and their topic modeling can classify the known to unknown types (Alpaydin, [2014](#)). It requires training sets and testing tests for the base of the model, so it can be classified as the case of an unknown type. Jockers and Mimno ([2013](#)) have used a supervised topic model to identify the gender of anonymous authors of 19th-century novels. These supervised topic models were derived from different programming software one of them was Voyant tools.

Voyant tools are free web based text analysis and visualization tools created by Stefen Sinclair, Geoffrey Rockwell, and their team. Voyant is a friendly used topic model with the ability to take extensive documentation, export data, with a simple user interface.

Conceptual Framework

This research adopts the conceptual framework of the Keyword Approach used by Binali et al. ([2010](#)) and the OCC model of emotion grounded in the Emotion Theory (Ortony et al., [1988](#)).

Emotion Theory

Emotion detection in web communication investigates psychological state, text style, and tone (Kim et al., [2004](#)). Emotion detection from the Emotion Theory in text aims to infer the emotions that influence the writing. These emotions can be happy, sad, and angry and these emotions are outwards in the things people say or write (Prinz, [2004](#)). This research looks for the emotions that are expressed through digital text.

Computational Approach for Emotions Detection

There are three domains in the emotion detection approach that comes under computational task, namely, keyword base, learning base and hybrid based approach (Binali et al., [2010](#)). These make use of syntactic, (e.g., n-gram, tags, phase pattern) and semantics datasets, (synonyms etc.) to extract emotions. These approaches take an effective lexicon as a knowledge rich linguistics resource to provide the general meaning of words (Binali et al.,

[2010](#)). The keyword based approach of emotion detection align suitably to conduct the current analysis; therefore, this study incorporates this approach as one of the dimension of the computational approach.

Keyword Based Approach

This approach depends on the presence of keywords, which also involves pre-processing with an emotion dictionary. Keyword approach is easy to implement because it includes identifying words to be searched in the selected text (Binali et al., [2010](#)).

Keyword Based Component:

- Data is collected in a corpus.
- The texts identify tokens (words) and sentences separately.
- Syntactic and semantic data is used for post-processing tasks before machine learning takes place.

OCC (Ortony/Clore/Collins) Model of Emotion

The OCC model of emotions looks for specifications of different kinds of emotions to classify the underlying emotion type (Ortony et al., [1988](#)). The emotions here are expressed by an agent (Prendinger & Ishizuka, [2005](#)). It includes 22 emotion categories that are designed to model human intelligence of emotions. The overall tokens of the model are, namely happiness, sadness, anger, disgust, surprise, and fear. The OCC types are, namely, Happy-for, Sorry-for, Resentment, Gloating, Hope, Fear, Relief, Disappointment, Pride, Self-Reproach, Appreciation, Gratitude, Anger, Gratification, Remorse, Liking, Disliking, Shame, Admiration, and Pity. Emotion theory is not the theory of language but emotions, thus it doesn't study emotions based on the specific word but a class that underlies the particular emotion type collectively.

Research Methodology

This research has used the following research methodology:

Computer-Assisted Textual Analysis

The development of new software is providing researchers with a combined qualitative and quantitative approach to analyze any text. This combination is also known as mixed methods. It extracts the strength and eliminates the weakness of both approaches (Brier & Hopp, [2011](#)). It

considers both way processes to present the results. The computer-assisted analysis increases the software and hardware skills of the researcher. It also allows the researcher to analyze the text with more rigour, scope, and size. The important aspects of both methods are presented below (Smith et al., [2015](#)).

Mixed Approach or Computational Approach

This method also known as the mixed method approach works as a bridge for both the qualitative and quantitative researches (Wiedemann, [2013](#)). The quantitative part here looks up keywords, the semantic frequency and the different graphic visualization of the text with numeric counting of the text for common themes (Roberts, [2000](#)). It deducts the important material and generalizes the data with codes and states the finding for the hypothesis (Smith et al., [2015](#)). However, the qualitative part takes symbols, common threads, the meaning of the words, summaries and contexts (McKee, [2003](#)). This deals with the interpretation and description of the data and induces the analysis on the bases of all the influential social factors (Deetz, [1977](#)). The mixed method approach takes a supervised software program that exists for both approaches and undertakes a theoretical framework, which can provide analysis both ways in one method (Smith et al., [2015](#)). A supervised software program for this study is the Voyant computational tools.

Tool for Textual Analysis

Voyant tools are user friendly sources that are employable for both beginner and advanced users. It is a recommended tool for scholars who want to determine the frequency and context of specific words in relation to the larger text (Welsh, [2014](#)). The present study takes five basic Voyant tools that include:

- *Cirrus*: word cloud showing frequent terms.
- *Reader*: a corpus that can be scrolled
- *Trends*: a distributed graph that shows terms.
- *Summary*: a simple textual over-view
- *Context*: a concordance that shows the occurrence of key words within the context (Boulton, [2016](#)).

Each tool is visualized and analyzed accordingly. The sub tools alongside these basic tools, like, terms, links, correlation, term berry, and phrases are briefed to answer the research question.

Textual Material for Computational Analysis

The study has selected the tweets related to Kashmir's recent issues covering around 4357 words for sentiment analysis under the theoretical perspective of emotion detection. Noticeably, these tweets were from around the world wherein people have expressed their emotions and disquiets concerning the current political event that took place in the recent past. The tweets were converted to corpus text, and collectively 15--word pages were uploaded to the Voyant tool for the analysis. The analytical section considered the sentence level of the corpus to analyze whether the text expresses negative/positive or neutral sentiments. The N-gram method of sentiment analysis has looked for strength lexicons, which purposed to look for tokens in the speech, which had different had types of emotions.

Analysis and Visualization

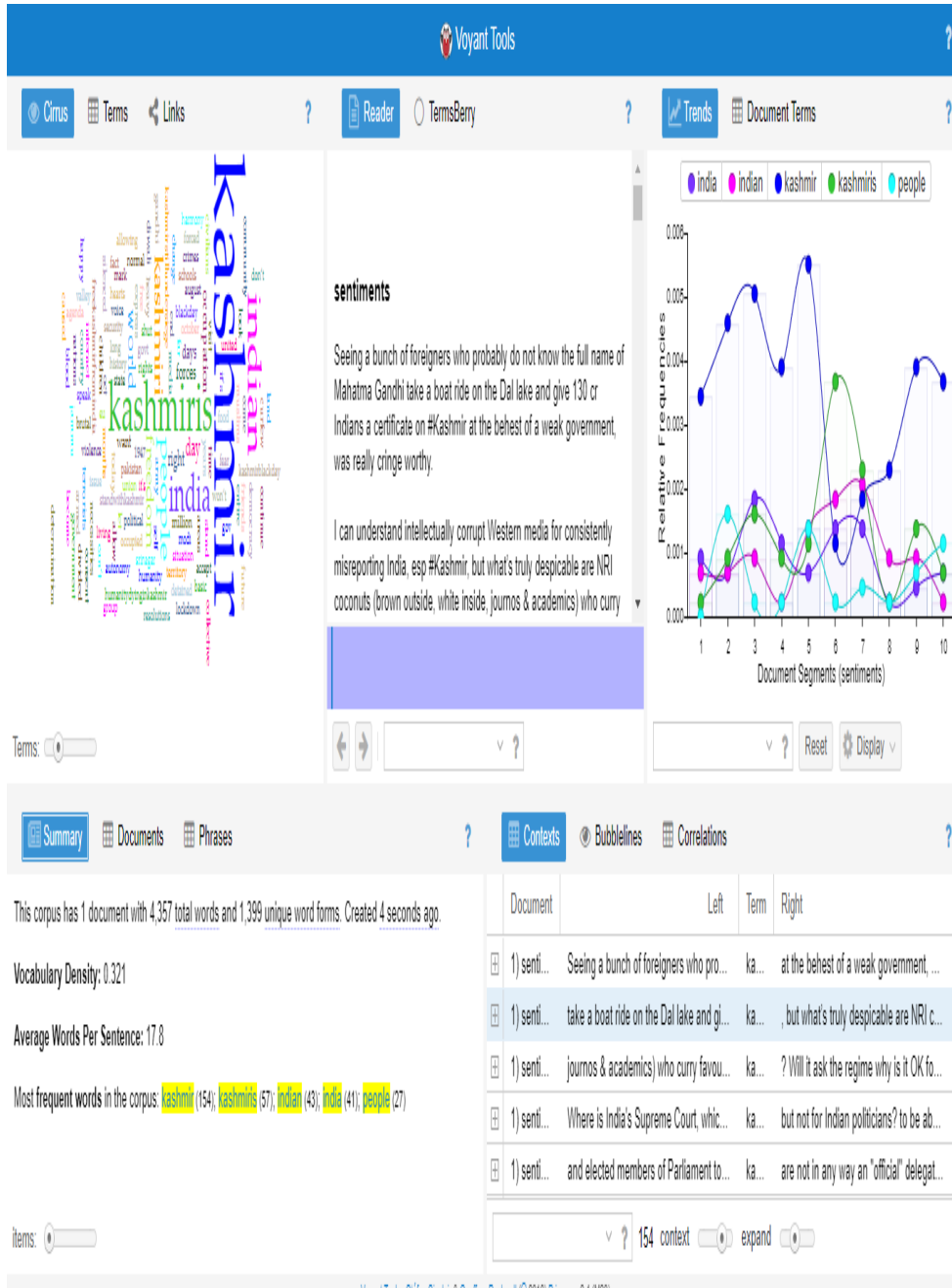
This section analyzes the collected text with five beginner tools. The sub tools alongside the main five tools are briefed to add up validity to the research questions and purpose. Each tool is explained and analyzed under the overall structure of the study. The visual of the first look is presented below just after the corpus is uploaded:

The apparent five dimensions are taken for the computational textual analysis. Each tool looked for sentiments suggesting that it is either inclined toward the quantitative or qualitative approach.

Cirrus

This tool is the first apparent skin, which displays the most frequent words used in the corpus with a cloud. It has its own item frequency option that determines; whether one wants to look into the broader words, deeper or less frequent words.

Figure 1
Visual Tool



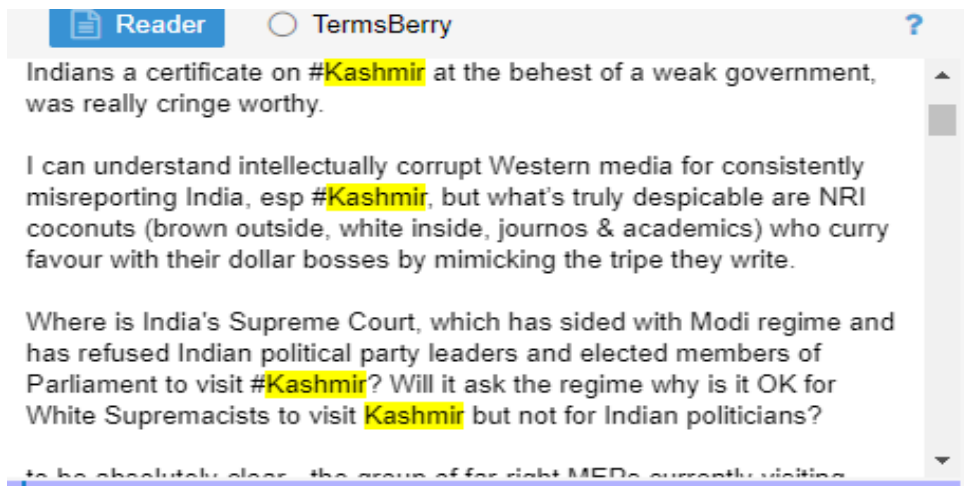
India and the Indian government. Different people around the world have expressed their sentiments with an epigraph stance that “we should all stand with Kashmir and resolve the issue that is under construction from 1947 onwards”. This tool cover-up the qualitative aspect of the research.

Reader

This segment of the tool presents the whole corpus file with a scrolling option aside. This tool is open to the researchers’ interpretation of social ideologies and hidden meanings. The researchers are required to read and re-read this segment of the tool for coming up with a valid description.

Figure 3

Reader Terms Berry



Most of the tweets have raised a question about the Modi regime and the UN, Union Territory status related to Jammu Kashmir. The tweets have used hashtags # with #FreeKashmir, FreeKashmirfromIndia, #BlackdayinKashmir #Kashmirstilundersiege #HumanityforKashmir, #stopcrefewinkashmir and #humanitydayinginKashmir, which have expressed the sentiment of people around the world for Kashmir and Kashmiri’s. This explains the phrase lexicon of N-gram. The overall sentiments are neutral and in favor of Kashmir. The overall corpus has pointed towards India and Indians for their brutality. The ideological perception of the tweets also revealed that Kashmir never got any support from the international powers of the political world like America and NETO besides knowing that the people of Kashmir are innocent and don’t deserve

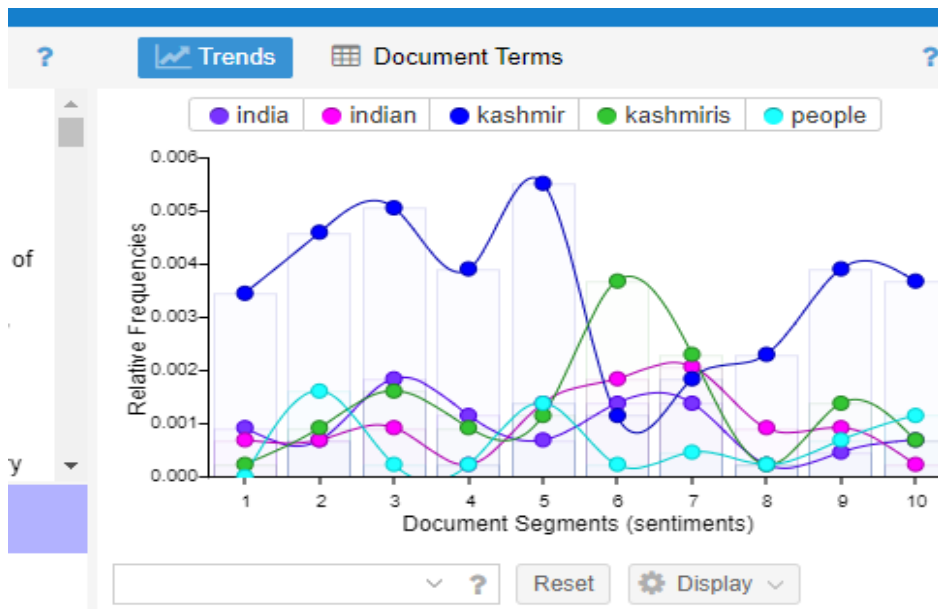
to be marginalized by the Indian government. This tool is more inclined towards the qualitative approach.

Trend

This skin shows the graph of the most frequent words in the document with their distribution. It shows the relative frequency of words that are used in almost all the tweets. This graph also tells the number of times the words are used alongside their intensity.

Figure 4

Trend



The graph shows 4 colours and each colour specifies a word. For example: Purple for India, Pink for India, Blue for Kashmir, Green for Kashmiris, and neon sea blue for its people. The N-gram is lexicon strength. There are 10 segments to the graph with blue showing with a higher frequency of 0.005 and neon sea blue with a lower frequency of 0.001. The green is the second most frequent word in the analysis with 0.003 and pink with 0.002 frequency, respectively. The overall distribution of the graph shows that tweets have a high level of neutral sentiments for Kashmir and Kashmiris and negative sentiments for India and Indians. The graph also interprets that the lower frequency (India and Indians) has a powerful

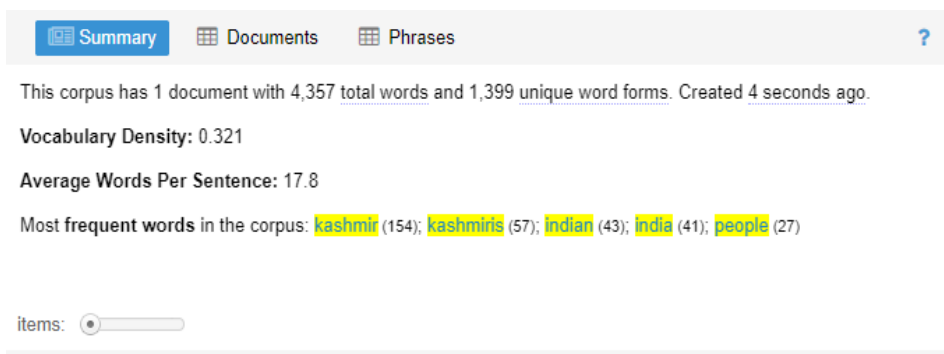
influence over the higher frequency (Kashmir and Kashmiris). The overall tool is more inclined towards a quantitative statistical approach.

Summary

This skin tells about the number of overall words, unique words, density and the average of all corpus.

Figure 5

Summary



The skin shows that with 4,357 words the vocabulary density is 0.321. This could have been 0.001 if the total words were around 7000. If the density decreases by 0.5, it would simultaneously raise doubts about the intensity of the uploaded corpus. The over vocabulary density with nouns, adjectives, verbs, and adverbs is good at 0.321, respectively. The unique words are 1,399, which are the most repeated and most frequently occurring words in the analysis. Those 1,399 words are the N-grams with; Kashmir (154), Kashmiris (57), Indian (43), India (41), and people (27). The same word distribution was presented in trend skin with graphs showing the relative frequency of these words. Twitter only allows 140 characters to be used one time and with that limit of 17.8% sentence average signifies the good intensity of the unique words. It also signifies the intellectual mindset of the people who were tweeting. This tool is more inclined towards a quantitative approach.

Context

The context skin is somewhat similar to the reading skin. This section deals with the context of a word or a set of words with keywords that have occurred time to time repeatedly.

Figure 6*Context*

Document	Left	Term	Right
1) senti...	Seeing a bunch of foreigners who pro...	ka...	at the behest of a weak government, ...
1) senti...	take a boat ride on the Dal lake and gi...	ka...	, but what's truly despicable are NRI c...
1) senti...	journos & academics) who curry favou...	ka...	? Will it ask the regime why is it OK fo...
1) senti...	Where is India's Supreme Court, whic...	ka...	but not for Indian politicians? to be ab...
1) senti...	and elected members of Parliament to...	ka...	are not in any way an "official" delegat...

154 context expand

This skin has the scrolling option and to whatever line you click it takes to the line number in the reading section. It has the option of showing sentences on the left and right with similar terms used in the corpus. It works the same as the reader option. It is open to the researchers' interruption of the text with reading, re-reading, and coding of the text. The term 'Kashmir' has occupied all the text; thus, the phrases on the left and right work in the context of this term.

Other Tools

The nature of the research questions is exploratory and allows us to take a view of the other beginner tools that helped in textual analysis. The other beginner tools include:

Terms

It shows the number of terms with the count of those words alongside. It is used after Cirrus to make the visualization simple and straight. It is inclined towards the quantitative approach. The N-gram represents all the words in the term section.

Figure 7
Terms

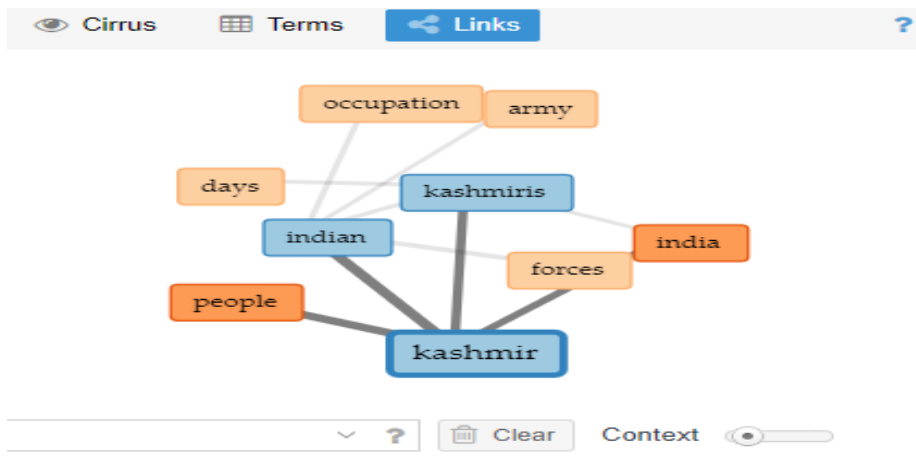
		Cirrus		Terms	Links	?
			Term	Count	Trend	
☰	<input type="checkbox"/>	1	kashmir	154		
☰	<input type="checkbox"/>	2	kashmiris	57		
☰	<input type="checkbox"/>	3	indian	43		
☰	<input type="checkbox"/>	4	india	41		
☰	<input type="checkbox"/>	5	people	27		
☰	<input type="checkbox"/>	6	kashmiri	23		
☰	<input type="checkbox"/>	7	world	19		
☰	<input type="checkbox"/>	8	freedom	18		
☰	<input type="checkbox"/>	9	day	15		
☰	<input type="checkbox"/>	10	right	15		

Link

It shows the link of 5 major words, Kashmir, Kashmiris, India, Indian and people. These five terms are shown interlinked with the whole corpus. These links are the N-grams of the corpus. It covers both the quantitative and qualitative parts.

Figure 8

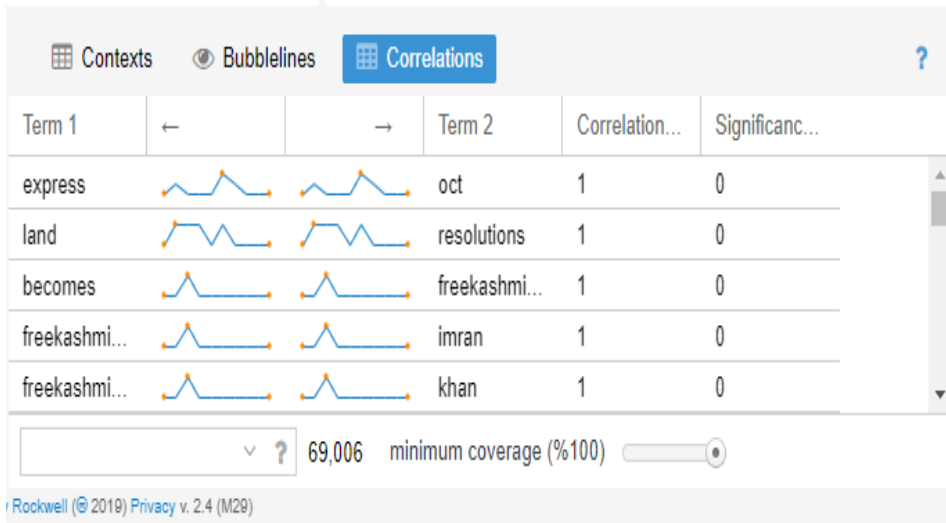
Link



Correlation

It shows the relation of the variables that were influential. The entire text was bounded by 1 and 1=100%. Thus, the text was co-related with 100 % frequency, respectively. The *N*-grams are all the tokens in the term section. This tool has an inclination towards quantitative approach.

Figure 9
Correlation



Term Berry

It shows the keywords in the linked balloons. The word is more frequently used with bigger balloons and the word is less frequently used with smaller balloons. It works the same as Cirrus with different visuals. The *N*-grams are all the words in the balloons. This segment covered both approaches.

Phrases

It comes alongside summary option. These phrases show the count of phrases per line and the length of a line in a corpus. It worked just like a summary, which shows the vocabulary, density, and unique words. All the phrases are *N*-grams. This part has an inclination towards a qualitative approach.

Findings and Discussion

This research presented five (5) basic Voyant tools with sub-tools that helped in computational textual analysis. Voyant worked as a supervised software program. The findings of the analysis are listed as follows:

- Voyant helped in grasping the broader context for interpretation. It allowed the researcher to understand the key elements of the intense corpus of 4357 words.
- The researchers could easily analyze the sentiments of people around the world for Kashmir and Kashmiri people. They expressed their concern by tweeting in hashtags, like, #FreeKashmir, #StandwithKashmir.
- They had also raised questions on the Union territory status for Jammu Kashmir. Globally, people had shown disappointment in the Modi regime and the Indian government for their brutality and viciousness.
- The word Kashmir with hashtag was used 174 times. The N-gram method and sentiment strength lexicon had identified that the sentiments of the people were neutral regarding Kashmir and Kashmiris, which had a visibly negative opinion of Indians.
- The overall sentimental analysis showed that people were neutral towards Kashmir's freedom. They were asking to end the 84 day long curfew in Kashmir and let the people of that region get back to their occupation and day-to-day life.
- The cirrus, trend, and summary skills showed results for the quantitative part; whereas the reader and context showed results for the qualitative part thus, the mixed methods approach was utilized in the analysis portion for each skin.
- Other beginner tools, terms, correlation, and term berry were supporting the quantitative approach, phrases, and links were inclined towards the qualitative part.
- The related tools showed the same results as the five beginner basic tools.

- Computational textual analysis is easy to interpret the meaning of the text comparatively than to do that manually. It saved time of the researchers and presented more accurate results for them.
- The computational analysis helped in both linguistics and literature textual interruption.

There are some pros and cons of the Voyant tools as far as the computer-assisted research method is concerned. These methods are mentioned below:

- The tools were easy to use.
- It was of no cost.
- It saved a lot of time as it presented everything visually in a few seconds.
- All the sub-skins were showing the same visual as the basic five skins.
- The tools were more inclined towards the quantitative method than the qualitative.
- The overall findings suggested that the software programs for textual analysis provided more accurate results.
- It was less biased and shows results from different dimensions.

Conclusion

The current study aimed to fill the existing gap in the field of digital humanities. It intended to explore a supervised model, which is easy and friendly to use by researchers and scholars of humanities and social sciences for textual analysis. Moreover, the textual analysis aimed to extract emotions from the selected tweets concerning the recent (2019) Kashmir issue. The theoretical perspective was followed accordingly in the analysis. Voyant was used for the computational analysis and basic beginner tools of Voyant were taken for testing with its description and analysis. All the tools helped in extracting the sentiment from the tweets of 4357 words. The opinions and feelings were neutral to negative. It was neutral for the freedom of Kashmir and negative towards the Indian government. The study also utilized computational textual analysis as a mixed-method approach. Therefore, the findings supported Voyant tools as an effective software for the textual analysis. The findings of this study are based on the answers to the research questions through analysis. Analysis has validated

that computer-assisted textual analysis is much more effective for the larger texts. The analysis further verifies Voyant as a supportive computational textual tool for analyzing human sentiments. The sentence level and N-gram sentiment analysis projected negative feelings against the Indians, while neutral feelings for the Kashmiris.

References

- Alpaydin, E. (2014). *Introduction to machine learning*. Massachusetts Institute of Technology.
- Barbosa, L., & Feng, J. (2010). *Robust sentiment detection on Twitter from biased and noisy data* [Poster presentation]. 23rd International Conference on Computational Linguistics: Posters (2009). Beijing. <https://aclanthology.org/C10-2005.pdf>
- Binali, H., Wu, C., & Potdar, V. (2010, April 13–16). *Computational approaches for emotion detection in text* [Paper presentation]. 4th International Conference on Digital Ecosystems and Technologies (IEEE 2009). Dubai, United Arab Emirates.
- Boulton, A. (2016). Integrating corpus tools and techniques in ESP courses. *ASP*, 69, 113–137. <https://doi.org/10.4000/asp.4826>
- Brier, A., & Hopp, B. (2011). Computer assisted text analysis in the social sciences. *Quality & Quantity*, 45, 103–128. <https://doi.org/10.1007/s1135-010-9350-8>
- Deetz, S. (1977). Interpretive research in communication: A hermeneutic foundation. *Journal of Communication Inquiry*, 31(1), 53–69. <https://doi.org/10.1007/BF00987104>
- Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: A comparative study on different approaches. *Procedia Computer Science*, 87, 44–49. <https://doi.org/10.1016/j.procs.2016.05.124>
- Evans, M. S. (2009). Defining the public, defining sociology: Hybrid science—Public relations and boundary-work in early American sociology. *Public Understanding of Science*, 18(1), 5–22. <https://doi.org/10.1177/0963662506071283>
- Gentzkow, M., Jesse M. S., & Matt, T. (2019). Measuring group differences in high-dimensional choices: Method and application to congressional

- speech. *Econometrica*, 87(4), 1307–1340. <https://doi.org/10.3982/ECTA16566>
- Go, A., Bhayani, R., & Huang, L. (2009). *Twitter sentiment classification using distant supervision*. CS224N Project. <https://tinyurl.com/mr3f99e2>
- Jockers, M. L., & Mimno, D. (2013). Significant themes in 19th-century literature. *Poetics*, 41(6), 750–769. <https://doi.org/10.1016/j.poetic.2013.08.005>
- Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing*, 42, 419–427. <https://doi.org/10.1007/BF02344719>
- MacKenzie, D., & Wajcman, J. (1999). *The social shaping of technology*. Open University Press.
- McKee, A. (2003). *Textual analysis: A beginner's guide*. Sage Publications Ltd.
- Mackay, H., & Gillespie, G. (1992). Extending the social shaping of technology approach: ideology and appropriation. *Social Studies of Science*, 22(4), 685–716. <https://doi.org/10.1177/030631292022004006>
- Mischaud, E. (2007). *Twitter: Expressions of the whole self. An investigation into user appropriation of a web-based communications platform* [Master thesis, University of London]. MEDIA@LSE. <https://www.lse.ac.uk/media-and-communications/assets/documents/research/msc-dissertations/2007/Mishaud-Final.pdf>
- Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press.
- Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining [Paper presentation]. The Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta.
- Pang, B., & Lee, L. (2004). *A sentimental education: Sentiment analysis using subjectivity analysis using subjectivity summarization based on*

- minimum cuts* [Paper presentation]. 42nd Association for Computational Linguistics. Barcelona, Spain.
- Pisarski, M. (2017). Digital postmodernism. *World Literature Studies*, 9(3), 41–53.
- Pontin, J. (2007, March 25). Artificial intelligence, with help from the humans. *New York Times*. <https://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html>
- Prendinger, H., & Ishizuka, M. (2005). The empathic companion: A character-based interface that addresses users' affective states. *Applied Artificial Intelligence*, 19(3-4), 267–285. <https://doi.org/10.1080/08839510590910174>
- Prinz, J. J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford University Press
- Roberts, C. W. (2000). A conceptual framework for quantitative text analysis. *Quality and Quantity*, 34, 259–274. <https://doi.org/10.1023/A:1004780007748>
- Saif, H., He, Y., & Alani, H. (2012, November 11–15). Semantic sentiment analysis of Twitter [Paper presentation]. 11th International Semantic Web Conference, Boston, MA, USA.
- Smith, J. A., Lloyd, M., & Pickard, V. (2015). COMPASS communication in action: Bridging research and policy. *Introduction. International Journal of Communication*, 9, 3411–3413.
- Welsh, M. E. (2014). Review of Voyant tools. *Collaborative Librarianship*, 6(2), 96–98.
- Wiedemann, G. (2013). Opening up to big data: Computer-assisted analysis of textual data in social sciences. *Historical Social Research/Historische Sozialforschung*, 38(4), 332–357.
- Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). *Recognizing contextual polarity in phrase-level sentiment analysis* [Paper presentation]. Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. Vancour.