



Research Article

Library Chat Transcript Evaluation for User Sentiment During the COVID-19 Pandemic

Kathryn Barrett
Liaison Librarian
University of Toronto Scarborough Library
Scarborough, Ontario, Canada
Email: kathryn.barrett@utoronto.ca

Ansh Sharma
Computer Science Student
University of Toronto Scarborough
Scarborough, Ontario, Canada
Email: ansh.sharma@alumni.utoronto.ca

Received: 1 Oct. 2024

Accepted: 17 Mar. 2025

© 2025 Barrett and Sharma. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 4.0 International (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

DOI: [10.18438/ebli30642](https://doi.org/10.18438/ebli30642)

Abstract

Objective – The purpose of this research was to explore user sentiment on Ask a Librarian, a consortial chat service for university libraries in Ontario, Canada, between 2019 to 2021. We tested how the characteristics of the chat (such as year, semester, user type, operator type, affiliation mismatch, and user complaints) and the onset of the COVID-19 pandemic affected sentiment scores.

Methods – The researchers analyzed 3,339 chat transcripts using VADER, a free, open-source Python natural language processing library for sentiment analysis. We tested the significance of relationships between study variables and sentiment score using either a two-samples t-test or ANOVA.

Results – Between 2019 to 2021, overall sentiment on Ask a Librarian was positive and higher among operators than users. There was a significant relationship between sentiment scores and operator type, affiliation mismatch, and complaints respectively. The year, semester, and pandemic status of the chat were also significantly associated with sentiment score. Chats that took place during the COVID-19 pandemic had a significantly higher overall sentiment score than pre-pandemic chats. Average user sentiment score was also higher during the pandemic, but there were no significant differences in average operator sentiment score.

Conclusion – The COVID-19 pandemic had a significant effect on the emotional tone of the overall chat interaction, as well as the sentiment within the user's messages. Practitioners can replicate our approach to understand user emotions, opinions, attitudes, or appraisals during times of disruption or emergency, as well as for regular service assessment.

Introduction

With the onset of the COVID-19 pandemic in March of 2020, academic libraries experienced an immediate and significant disruption to their operations. As colleges and universities across North America closed their physical spaces and shifted courses online, academic libraries switched to online service delivery, including launching or expanding virtual reference services (Radford et al., 2020; Yaticilla & Young, 2021). Chat reference services were particularly well-positioned to play a role in pandemic response. Live chat offers synchronous assistance to users in the online environment, and it meets user preferences for convenience, efficiency, and personal and informal communication styles (Chow & Croxton, 2014; Connaway & Radford, 2011; Mawhinney, 2020). Unsurprisingly, many chat services saw surges in demand and rises in chat volume throughout the pandemic (Radford et al., 2022; Osorio & Droog, 2021).

While some academic libraries launched new online reference services to respond to the pandemic (Decker & Chapman, 2022), most already had a virtual reference service in place prior to COVID-19 (Cohn & Hyams, 2021; Osorio & Droog, 2021). For example, a 2018 survey of ARL libraries found that 91% offered some form of virtual reference (Catalano et al., 2018). Consequently, the pandemic transition for most libraries involved improving access to existing virtual reference services. Common strategies included training new chat operators, increasing shifts, expanding service hours, drafting best practice documents, creating new triaging workflows, implementing new features for the chat software, and making the chat service more prominent (Cohn & Hyams, 2021; Murphy et al., 2022; Osorio & Droog, 2021).

Researchers are beginning to explore how the COVID-19 pandemic changed the nature of chat reference interactions themselves, such as volume, temporal distribution, duration, type, complexity, instructional content, number of complaints, and relational aspects of chat questions (Barrett et al., 2024; De Groote & Scoulas, 2021; Hervieux, 2021; Munip et al., 2022; Radford et al., 2022; Watson, 2023). For example, Radford et al. (2022) described shifting levels of deference, including politeness and expressions of gratitude and frustration, in user messages during the pandemic. Our study aims to contribute to this literature by extending our understanding of how the pandemic affected the emotional tenor of chat interactions. We conducted a sentiment analysis of chat transcripts from a large, consortial chat service in Ontario, Canada, and compared chats from the pre-pandemic period in 2019 to pandemic-era chats from 2020 and 2021.

Literature Review

An understanding of user needs enables librarians to provide efficient and accurate reference services. Reviewing activity on the library's various reference services can help staff to identify common patron needs. Given that chat reference generates and preserves a large volume of data in the form of chat records and transcripts, librarians can review this data to identify common user needs and ensure that chat personnel can receive appropriate training to provide high quality service (Wang, 2022). Historically, researchers have used qualitative methods to identify chat trends, such as hand-coding chat transcripts, but these methods are time-consuming and ill-suited to the large datasets generated by live chat (Chen & Wang, 2019). Consequently, researchers are beginning to explore automated, computational approaches to analysis, such as text mining and machine learning, often using natural language processing techniques (Kohler, 2020; Paulus et al., 2019).

Several researchers have conducted studies to explore automated methods for the topical analysis of virtual reference records. For example, Brousseau et al. (2021) used a supervised machine learning model to code transcripts, and Turp & Hervieux (2023) used regular expressions to identify themes in virtual reference. One common approach in the literature is topic modeling, a natural language processing technique that reveals the hidden structure within documents by grouping words with similar meanings and separating words with different meanings (George & Birla, 2018). Several researchers have conducted studies to explore the viability and application of different topic modeling techniques to chat reference data. For example, Ozeran and Martin (2019) tested different algorithms for topic modelling and determined that Latent Dirichlet Allocation, Phrase-Latent Dirichlet Allocation, and Non-Negative Matrix Factorization were the most promising for large datasets. Koh and Fienup (2021) qualitatively measured the accuracy and interpretability of different topic modelling techniques and judged that Probabilistic Latent Semantic Analysis performed the best. Sharma et al. (2022) incorporated a mix of targeted searching for query terms using regular expressions and natural language processing using the spaCy library and found that it was effective for topical analysis of chat transcripts.

Other researchers have applied topic modelling techniques to learn about aspects of their chat services. Schiller (2016) explored the learning taking place on Wright State University's chat reference service using a mix of manual and automated coding using a text mining software, finding that two teaching styles, "give fish" and "teach fishing," are constructed in the process of mediated learning within the chat interaction, which is facilitated by the chat technology and the social environment. Kohler (2017) used topic extraction algorithms to identify popular chat topics, with the results showing that general help, database searching, interlibrary loan requests, catalogue searching, and login information were common topics. Walker and Coleman (2021) predicted the difficulty of incoming chat questions using machine learning and natural language processing techniques, and found that the predictive power of the modeling processes was statistically significant. Recently, researchers have also used topic modeling to understand how the COVID-19 pandemic affected the nature of chat topics, finding that the content of questions remained largely unchanged (Sobol et al., 2023).

Another popular computational approach employing natural language processing is sentiment analysis. Also known as opinion analysis or opinion mining, sentiment analysis extracts patterns of information from textual data based on the author's emotions, such as their thoughts, attitudes, views, opinions, beliefs, or preferences (Lamba & Madhusudhan, 2022). Sentiment analysis extracts feelings in the form of polarity, measured on a scale of -1 (very negative) to +1 (very positive), with 0 representing neutrality (Lamba & Madhusudhan, 2022). Sentiment analysis has many applications in business, because it can be applied to customer reviews to detect changes in client opinion and improve customer support (Liu et al.,

2020). Within libraries, it can be applied to data from patron feedback, reference transactions, and social media to provide insights about user satisfaction (Lamba & Madhusudhan, 2022).

There is a small but growing body of literature about sentiment analysis in libraries. While one study reported on a sentiment analysis of library tweets (Lund, 2020), the majority of sentiment analysis research has examined chat transcripts. For example, Kohler (2017) found that sentiment was overwhelmingly positive on Greenlease Library's chat service, while Brousseau et al. (2021) determined that the number of satisfied chats at Brigham Young University Library decreased over a three-year period. Several recent studies have looked at the impact of the COVID-19 pandemic on user sentiment, with mixed results. Kathuria (2021) conducted sentiment analysis on Georgia State University's chat transcripts from 2019 to 2020, finding that overall sentiment was much lower during the pandemic. There was a spike in positive words early during the COVID-19 pandemic, but sentiment dropped during summer and fall of 2020. Kohler (2020) used the VADER sentiment analysis tool to evaluate chat transcripts from 2020 at Virginia Tech and found that sentiment scores were overwhelmingly positive, with the small group of negative chats mainly being cases of an inherently negative research topic or lack of access to specific resources. Sobol et al. (2023) used the Linguistic Inquiry and Word Count tool for sentiment analysis of transcripts from a consortial chat service covering 2019-2020 and 2020-2021. Overall, the emotional tone of chats was positive, and higher in the messages of patrons than providers. During the pandemic, the positive language of chat providers declined, while sentiment scores for patrons had a small increase.

Aims

The aim of this research was to explore user sentiment on the Ask a Librarian chat service between 2019 to 2021, with a particular focus on how the characteristics of the chat and the onset of the COVID-19 pandemic affected sentiment scores. We sought to answer the following research questions:

1. What is the average sentiment score on Ask a Librarian?
2. Do average user and operator sentiment scores differ?
3. Are there significant differences in sentiment score based on user or operator type?
4. Does an affiliation mismatch between the user and operator affect sentiment scores?
5. Did sentiment scores vary by year or semester?
6. Was there a significance difference between pre-pandemic and pandemic sentiment scores?
7. How does the presence of a complaint in the chat transcript affect sentiment scores?

Methods

Background and Setting

Scholars Portal is the digital services arm of the Ontario Council of University Libraries (OCUL), a consortium representing the libraries of the 21 universities in the province of Ontario, Canada. Scholars Portal manages Ask a Librarian, a collaborative chat service offering real-time library- and research-related assistance from librarians, paraprofessional library staff, and graduate student employees. The service is offered at 16 participating universities for 67 hours per week during the academic year, reaching approximately 445,000 full-time equivalent students, and receiving over 25,000 chats a year. The researchers received approval for this study from the Research Ethics Board of the University of Toronto, the home institution of the authors, in addition to Scholars Portal's Ask a Librarian Research

Data Working Group. Users are informed that their chat data can be used for research purposes through Ask a Librarian's privacy policy, and operators are informed during training.

Data Collection, Sampling, and Preparation

This research study employed two approaches to transcript analysis: manual coding for select variables and natural language processing for sentiment analysis. Manual coding was performed to enable us to determine if characteristics of the chat interaction were associated with the chat's sentiment score. For hand-coding to be achievable for the research team, we selected a sample of chats rather than analyzing the entire corpus from the study period.

All English-language chats that took place between January 1, 2019, to December 31, 2021, were eligible for sampling. This study excluded French-language chats (due to the language skills of the research team) and text message (SMS) interactions. In total, 124,080 eligible chats occurred over this period.

The researchers downloaded a metadata spreadsheet for the eligible chats from LibraryH3lp, the chat software. After removing identifying information about the user and operator, we created new variables in the spreadsheet to record the year and the semester that the chat took place. We operationalized the winter semester as the months of January – April, summer as May – August, and fall as September – December. Through this process, each chat was assigned to one of 9 possible semesters from the study period.

To create samples for each of the 9 semesters, we used Excel to randomly select chats according to their unique ID in the metadata spreadsheet. Sample sizes were calculated for each semester to achieve a 95% confidence level. Overall, we selected 3,339 chats from the 9 semesters across the three-year study period.

Variable Creation and Coding

To determine whether each chat took place before or during the pandemic, we created a variable in the metadata spreadsheet to record whether the chat took place before or after the World Health Organization declared COVID-19 a pandemic. Pre-pandemic chats occurred on March 10, 2020, or earlier. Pandemic-era chats occurred on or after March 11, 2020.

One team member (KB) hand-coded two additional variables by reviewing the complete transcript of each sampled chat:

1. **User type:** This variable referred to the user's status at the university. It was coded based on the user's response to an auto-generated prompt at the beginning of the chat requesting that they share information about themselves. The options were: undergraduate student, graduate student, faculty member, staff member, alumni, member of the public, or other. If the user did not respond to the prompt, their type was recorded as unknown.
2. **Complaint:** This variable recorded whether there was at least one complaint present within the chat transcript, which we defined as any expression of grievance, dissatisfaction, injustice, or wrong suffered on the part of the patron. This could be any statement from the user that something had gone wrong, was not good enough, was unsatisfactory, or was unacceptable. Given the subjectivity of identifying complaints, we chose to be inclusive and coded problems encountered by users as complaints.

Sentiment Analysis

We used VADER (Hutto & Gilbert, 2014), a Python natural language processing library, to analyze the chat transcripts. VADER is a simple rule-based model for general sentiment analysis. It can be used for text across domains, but it performs especially well in the analysis of social media text. We selected VADER because it is a free and open source tool, and because it is especially attuned to sentiments expressed in social media, which made it a good fit for our corpus of online chat data.

The VADER library processed a .csv file made up of rows for each chat, with columns containing the metadata fields and corresponding transcript. The text within each transcript was analyzed by parsing every message within the interaction and assigning each message a score.

The toolchain distinguished whether a particular message was sent by the user or the operator through the content of the message in the chat transcript. Messages beginning with the system-generated operator tag (automatically included in LibraryH3lp chat transcripts) were assumed to be sent from the operator. Messages beginning with the guest identification string (automatically assigned by the LibraryH3lp platform) were assumed to be sent from the user.

The toolchain processed the data and exported a .csv spreadsheet with its output. Identifying data was automatically removed from the spreadsheet by the toolchain, including metadata fields related to the user and operator, as well as the complete text of the transcript.

The output spreadsheet added several new fields for sentiment score:

1. **Average VADER user sentiment score:** mean sentiment score calculated by VADER for all messages sent by the user within the chat transcript
2. **Average VADER operator sentiment score:** mean sentiment score calculated by VADER for all messages sent by the operator within the chat transcript
3. **Average VADER overall sentiment score:** calculated by the researchers, the mean of the combined average user and operator sentiment scores, reflecting the overall sentiment across all the messages within the chat transcript

The toolchain also processed and recorded two additional variables in the output spreadsheet:

1. **Operator type:** This variable referred to the operator's position within the library and was determined based on the operator username in the chat metadata. The toolchain looked up the username in a spreadsheet containing each active operator's role at their home library and recorded the response in a new column. The options were: librarian, library technician, student employee, or unknown. In the Ask a Librarian context, librarians have graduate degrees in library or information science, technicians have a college diploma for library and information technicians (some may also have an advanced degree in LIS), and student employees are graduate students enrolled in a library or information science program who have received reference training.
2. **Affiliation mismatch:** This variable recorded whether the user and operator were affiliated with the same institution. The toolchain compared the queue through which the chat was submitted (the user's university) and the operator's username (which includes a suffix for their university) in the chat metadata. If they were affiliated with the same institution, the chat was recorded as an affiliation match. If they were not, it was recorded as an affiliation mismatch.

Data Compilation and Data Analysis

We merged the spreadsheets containing the chat metadata, the constructed and hand-coded variables, and the VADER output into a single spreadsheet based on unique chat ID.

In IBM SPSS Statistics, we generated descriptive statistics and tested the significance of the relationships between variables and sentiment score using two-sample t-tests and Analysis of Variance (ANOVA). A two-samples t-test compares the means of two groups to determine whether the associated population means are significantly different. ANOVA is a statistical test used to determine whether there is a significant difference in the means of more than two groups.

When interpreting results, we used typical threshold values for VADER to determine if sentiment scores were positive, neutral, or negative (Hutto, 2014):

1. Positive sentiment: ≥ 0.05
2. Neutral sentiment: < 0.05 and > -0.05
3. Negative sentiment: ≤ -0.05

Results

Average Sentiment Scores

The mean overall VADER sentiment score on Ask a Librarian between 2019 to 2021 was positive, $M = 0.213$. Average sentiment was higher for operators than users (see Table 1).

Table 1
Average Sentiment Scores on Ask a Librarian, 2019 to 2021

Sentiment Score	<i>M</i>	<i>SD</i>
Average Overall	0.213	0.120
Average Operator	0.236	0.175
Average User	0.195	0.150

Association Between User and Operator Type and Sentiment Score

An ANOVA test showed that user type was not significantly associated with average overall sentiment scores ($p = .498$). User type approached but did not meet significance for average user sentiment scores ($p = .059$) and for average operator sentiment scores ($p = 0.06$). For details, see Table 2.

Table 2
 AVOVA for Sentiment Score and User Type

Sentiment Score	Undergrad		Graduate		Faculty		Staff		Member of Public		Alumni		Unknown		Other		df	F
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD		
Average Overall	0.215	0.115	0.213	0.107	0.212	0.126	0.195	0.113	0.234	0.129	0.199	0.099	0.213	0.132	0.208	0.122	7, 2832	0.909
Average Operator	0.241	0.163	0.23	0.15	0.237	0.174	0.17	0.145	0.233	0.188	0.21	0.16	0.242	0.198	0.233	0.156	7, 3086	1.934
Average User	0.193	0.138	0.199	0.134	0.191	0.145	0.229	0.147	0.235	0.148	0.2	0.132	0.189	0.17	0.195	0.16	7, 3072	1.944

An ANOVA test determined that operator type was significantly associated with overall average sentiment scores ($p < .001$). Mean sentiment was lowest among the library technician group and highest among the student employee operator group. Operator type was not significantly associated with average user sentiment scores ($p = .972$), but it was significantly related to average operator sentiment scores ($p < .001$). Mean operator sentiment scores were lowest among library technicians and highest among the student employee group. For details, see Table 3.

Table 3
ANOVA for Sentiment Score and Operator Type

Sentiment Score	Librarian		Library Technician		Student Employee		Unknown		df	F
	M	SD	M	SD	M	SD	M	SD		
Average Overall	0.208	0.119	0.2	0.119	0.232	0.121	n/a	n/a	2, 2837	18.299***
Average Operator	0.217	0.166	0.208	0.157	0.268	0.158	0.284	0.279	3, 3090	29.957***
Average User	0.196	0.140	0.195	0.153	0.196	0.156	n/a	n/a	2, 3077	0.028

*** $p < 0.001$

Association Between Affiliation Mismatch and Sentiment Score

A two-samples t-test showed that mean VADER overall sentiment score was significantly lower in chats in which there was an affiliation mismatch between the user and the operator compared to chats in which the user and operator were from the same institution (see Table 4). In chats with affiliation mismatches, average sentiment scores were lower for both user messages and operator messages.

Table 4
Two-Samples T-Test for Sentiment Score and Affiliation Mismatch

Sentiment Score	Match		Mismatch		df	t	p
	M	SD	M	SD			
Average Overall	0.222	0.123	0.205	0.118	2826.852	3.742	< .001***
Average Operator	0.243	0.171	0.224	0.166	2958	3.019	0.003**
Average User	0.203	0.149	0.187	0.151	3071	2.968	0.003**

** $p < 0.01$

*** $p < 0.001$

Association Between Year, Semester, and Pandemic Status and Sentiment Score

An ANOVA test showed that the effect of year on overall average VADER sentiment score was significant ($p = 0.03$). Average sentiment score was lowest in 2019 and highest in 2020. The effect on average patron score and average operator score was not significant ($p = 0.122$ and $p = 0.505$ respectively). See Table 5 for details.

Table 5
ANOVA for Sentiment Score and Year

Sentiment Score	2019		2020		2021		df	F
	M	SD	M	SD	M	SD		
Average Overall	0.205	0.114	0.218	0.126	0.217	0.121	2, 2837	3.526*
Average Operator	0.232	0.175	0.234	0.176	0.241	0.174	2, 3091	0.684
Average User	0.188	0.146	0.2	0.15	0.198	0.153	2, 3077	2.102

* $p < 0.05$

An ANOVA test showed the effect of semester on overall average VADER sentiment score was significant ($p < .001$). The semesters with the highest average sentiment scores were summer 2021 and summer 2020. The semesters with the lowest average sentiment scores were summer 2019 and fall 2021. Additional ANOVA tests showed that the effect of semester was significant on average patron sentiment scores ($p = .01$), but not on average operator sentiment scores ($p = .103$). See Table 6 for details.

Table 6
 AVOVA for Sentiment Score and Semester

Sentiment Score	Winter 2019		Summer 2019		Fall 2019		Winter 2020		Summer 2020		Fall 2020		Winter 2021		Summer 2021		Fall 2021		df	F
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD		
Average Overall	0.215	0.119	0.19	0.1	0.209	0.118	0.217	0.127	0.228	0.134	0.21	0.118	0.21	0.129	0.234	0.122	0.206	0.11	8, 2831	3.461***
Average Operator	0.244	0.19	0.215	0.168	0.237	0.164	0.233	0.173	0.244	0.194	0.225	0.159	0.23	0.181	0.258	0.176	0.235	0.165	8, 3085	1.661
Average User	0.197	0.152	0.181	0.125	0.185	0.158	0.199	0.15	0.206	0.163	0.196	0.138	0.203	0.157	0.216	0.163	0.176	0.137	8, 3071	2.527**

** $p < 0.01$

*** $p < 0.001$

A two-samples t-test found that average overall sentiment scores and average user sentiment scores were higher during the pandemic compared to pre-pandemic. There was no significant difference in operators' average sentiment scores. See Table 7 for details.

Table 7
Two-Samples T-Test for Sentiment Score and Pandemic Status

Sentiment Score	Pre-Pandemic		Pandemic		df	<i>t</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Average Overall	0.206	0.115	0.218	0.124	2601.858	-2.574	0.01**
Average Operator	0.233	0.173	0.238	0.176	3092	-0.806	0.421
Average User	0.189	0.147	0.2	0.152	2679.729	-2.071	0.038*

* $p < 0.05$

** $p < 0.01$

Association Between Complaints and Sentiment Score

A two-samples t-test showed that average overall VADER sentiment scores were significantly lower when the user had at least one complaint compared to chats without complaints. In chats with complaints, average sentiment scores were significantly lower for both user messages and operator messages. See Table 8 for details.

Table 8
Two-Samples T-Test for Sentiment Score and Complaints

Sentiment Score	No Complaint		Complaint		df	<i>t</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Average Overall	0.222	0.12	0.176	0.115	2838	7.992	< .001***
Average Operator	0.242	0.176	0.207	0.168	3092	4.257	< .001***
Average User	0.206	0.152	0.147	0.133	916.119	9.374	< .001***

*** $p < 0.001$

Discussion

Chat Characteristics and Sentiment Score

Between 2019 to 2021, overall sentiment on the Ask a Librarian service was positive, with a mean sentiment score of 0.213. Sentiment differed between the participants of the chat; the average sentiment score of operators was higher than that of users. It is difficult to determine if this observation is valid,

because it is possible that the content of the user's research question or information need may confound sentiment score. For example, certain research topics may contain words that VADER assigns a negative score, which may contribute to user messages having a more negative sentiment score than the operator messages. It's worth noting that Sobol et al. (2023) found the opposite pattern. In their sample, patrons had higher positive language than chat providers.

We found no statistically significant relationships between user type and sentiment scores, meaning that there were no differences in sentiment between students, faculty members, staff, alumni, or members of the public. In contrast, we did identify a statistically significant association between operator type and average overall sentiment scores, as well as average operator sentiment scores. For both types of sentiment scores, we found the scores to be lowest when the operator was a library technician and highest when the operator was a student employee. Further research is needed to explore why this relationship exists.

Chats in which there was a mismatch in affiliation between the user and operator had significantly lower sentiment scores. This may be because, as previous research has shown, users are more likely to be dissatisfied when they are made aware that they are being assisted by a library staff member from outside of their home institution, as patrons may perceive these operators as lacking knowledge about their local context (Barrett & Pagotto, 2019, 2021). The pandemic also likely exacerbated the difficulty of serving patrons from other libraries, as the shifting conditions of pandemic-era services may have made it difficult to share information between libraries and efficiently and accurately answer users' questions. We also determined that chats containing at least one complaint had a lower overall sentiment score, user sentiment score, and operator sentiment score. While this association may seem obvious, and the test redundant, we tested the relationship between these variables to determine if VADER was being influenced by the tone or attitude of chat participants, given that the tone of complaints is inherently negative.

Temporal Aspects of Sentiment Score

There was a statistically significant relationship between the year the chat took place and average overall sentiment score. Surprisingly, mean scores were lowest in the pre-pandemic year, 2019, and highest during the year in which the pandemic began, 2020. This result reinforces those of Kathuria (2021) and Radford et al. (2022), whose studies noted more positive words and expressions near the onset of the pandemic in 2020. In addition, there was a statistically significant relationship between semester and overall sentiment score, with the lowest sentiment occurring pre-pandemic, in summer 2019, and the highest sentiment score occurring during the pandemic, in summer 2021. Regardless of relative differences in sentiment score between years and semesters, we note that the average sentiment scores always remained above the threshold of 0.05, reflecting positive sentiment.

Results about the significance of the year and semester of the chat are consistent with our findings related to pandemic status: chats that took place after the WHO declared COVID-19 a pandemic had a significantly higher overall sentiment score than pre-pandemic chats. Our results are consistent with Kohler's research (2020), which found that sentiment was positive during the pandemic, suggesting a "civility of discourse," and differ from those of Kathuria (2021), who identified an increase in negative sentiment during the pandemic. The differences in our results may be due to the nature of our samples: Kathuria drew on data from 2019 to 2020, Kohler from 2020 alone, and our research covered 2019 to 2021. There were important contextual differences in the pandemic and library services across these time periods that may have affected sentiment score. For example, throughout the pandemic, there were

periods of COVID surges and resulting lockdown or stay-at-home orders, which restricted the availability of library spaces and collections. Our sample captures Ontario's second state of emergency (beginning in January 2021), the third wave of the virus, rising infections from COVID variants of concern, and Ontario's third state of emergency (beginning in April 2021). The states of emergency triggered stay at home orders, which prompted Ontario's academic libraries to close and shut down services like curbside pickup and scan and deliver. Overall, 2021 was a period of significant flux, with users losing and gaining access to physical spaces and collections, which may have influenced sentiment scores differently than earlier phases of the pandemic. In addition, the variation in our results may also be due to the different sentiment analysis tools we used; Kathuria grouped words into positive or negative sentiment using a coding system, while Kohler, like us, used VADER.

The VADER library's ability to calculate average patron sentiment score and average operator score for each chat lead us to a noteworthy finding: average user sentiment score was higher during the pandemic, while there was no significant difference in average operator sentiment score. This indicates that it may have been the user's tone or attitude that contributed to statistically significant differences in sentiment score on Ask a Librarian during the pandemic. Sobol et al. (2023) noted a similar trend on their consortial chat service: scores for patron chats had a small increase in positive language during the pandemic, while positive language among chat providers declined. Additional research is needed to determine why user sentiment score increased during the pandemic. A study by Radford et al. (2022) may provide an initial explanation. Many chat operators reported positive changes in user communication style during the pandemic, such as politeness and expressions of gratitude. Kohler (2017) also noted the role of politeness in positive sentiment and added that the user's sense of being part of the same academic community as the operator may influence the language used. These elements of communication style may have been parsed positively by the VADER library.

Limitations and Future Research

Our study has several limitations. Firstly, there is a great deal of complexity and nuance in textual data, meaning that sentiment analysis tools may sometimes parse text more inaccurately than a human researcher. Qualitative studies of pandemic-era chat discourse would be a helpful complement to the existing computational studies. Secondly, because we employed a mix of hand-coding and natural language processing for our research, we selected a sample of chats rather than processing the entire corpus of chats between 2019 to 2021 using VADER. Although our sample size was large (>3000 chats), failing to utilize the entire population of chat transcripts may have limited the generalizability of our findings. Future research could employ our methods using a larger sample size of chats. In addition, given the nature of our dataset, many of our findings are contextual to the first two years of the pandemic. While our results provide a rich portrait of users' sentiments from 2020-2021, further research is needed to explore user sentiment during the later years of the pandemic (2022-2023). Additional studies are also needed to determine if the associations we uncovered between chat characteristics and sentiment score will extend beyond the pandemic. Finally, the VADER library is somewhat sensitive to the subject of the conversation, meaning that research topics containing negative terms may result in negative sentiment scores. Additional research could explore methods to effectively control for the subject of the chat.

Implications

Our research study outlines a methodology for chat transcript analysis that combines hand coding and natural language processing. This approach enables researchers to calculate sentiment scores for chat

transcripts using the VADER library and run inferential statistics to test the relationships between the hand-coded or toolchain-generated variables and sentiment scores. This allows for deeper investigations into how the characteristics of the chat affect sentiment score.

While we used this methodology to examine chat transcripts from the COVID-19 pandemic on a consortial chat service, this approach could be used to explore the nature of chat interactions over any period and on any type of chat service. VADER can also process content from other forms of virtual reference (such as emails) or social media. Practitioners can incorporate this methodology into regular service evaluation or review to understand user sentiment and satisfaction. Given that using sentiment analysis tools is less time- and labour-intensive than traditional hand-coding, sentiment analyses could be run more regularly for real-time assessment.

As libraries increasingly operate in environments of evidence based or data-informed decision-making, sentiment analysis can be a helpful approach to identify areas where libraries can make improvements to customer service, training, policies, or service models. As a free, open-source tool, VADER is ideal for librarians beginning to explore sentiment analysis.

Conclusion

This study reports on the sentiment analysis of over 3,000 chat transcripts from Ask a Librarian from 2019 to 2021. Overall, we found that mean sentiment was positive (>0.2), and higher among operators than patrons. This difference in the sentiment of participants may be due to the inherent negativity of some users' research topics or the problems they were describing. Several characteristics of the chat were significantly associated with sentiment scores, namely operator type, affiliation mismatch, and complaints. Sentiment score also varied significantly over time: it was lowest in 2019 and highest in 2020. The COVID-19 pandemic was also significant: chats that took place during the pandemic had a higher average overall sentiment score and higher average user sentiment score. The results of this study indicate that Ask a Librarian met user needs during the pandemic, as the polarity of sentiment scores remained positive during pandemic-related disruptions in library operations. We recommend that sentiment analysis continue to be conducted as part of regular virtual reference assessment.

Author Contributions

Kathryn Barrett: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Supervision, Writing – original draft, Writing – review & editing **Ansh Sharma:** Methodology, Data curation, Software, Writing – review & editing

Acknowledgements

The authors wish to thank Scholars Portal for providing access to the chat metadata and transcripts to support this project. The team also wishes to acknowledge Kirsta Stapelfeldt and Chad Crichton of the University of Toronto Scarborough Library for their contributions to the research project examining the Ask a Librarian service during the COVID-19 pandemic.

References

- Barrett, K., Crichton, C., & Logan, J. (2024). An analysis of user complaints on chat reference during the COVID-19 pandemic: Insights into user priorities. *Internet Reference Services Quarterly*. Advance online publication. <https://doi.org/10.1080/10875301.2024.2379816>
- Barrett, K., & Pagotto, S. (2019). Local users, consortial providers: Seeking points of dissatisfaction with a collaborative virtual reference service. *Evidence Based Library and Information Practice*, 14(4), 2-20. <https://doi.org/10.18438/ebliip29624>
- Barrett, K., & Pagotto, S. (2021). Pay (no) attention to the man behind the curtain: The effects of revealing institutional affiliation in a consortial chat service. *Partnership*, 16(2), 1-21. <https://doi.org/10.21083/partnership.v16i2.6651>
- Brousseau, C., Johnson, J., & Thacker, C. (2021). Machine learning based chat analysis. *Code4Lib*, 50.
- Catalano, A. J., Glasser, S., Caniano, L., Caniano, W., & Paretta, L. (2018). An analysis of academic libraries' participation in 21st century library trends. *Evidence Based Library and Information Practice*, 13(3), 4-16. <https://doi.org/10.18438/ebliip29450>
- Chen, X., & Wang, H. (2019). Automated chat transcript analysis using topic modeling for library reference services. *Proceedings of the Association for Information Science & Technology*, 56(1), 368-371. <https://doi.org/10.1002/pr2.31>
- Chow, A. S., & Croxton, R. A. (2014). Usability evaluation of academic virtual reference services. *College & Research Libraries*, 75(3), 309-361. <https://doi.org/10.5860/crl13-408>
- Cohn, S., & Hyams, R. (2021). Our year of remote reference: COVID19's impact on reference services and librarians. *Internet Reference Services Quarterly*, 25(4), 127-144. <https://doi.org/10.1080/10875301.2021.1978031>
- Connaway, L. S., & Radford, M. L. (2011). *Seeking synchronicity: Revelations and recommendations for virtual reference*. Online Computer Library Center, Inc. <https://www.oclc.org/research/publications/2011/synchronicity.html>
- De Groote, S., & Scoulas, J. M. (2021). Impact of COVID-19 on the use of the academic library. *Reference Services Review*, 49(3/4), 281-301. <https://doi.org/10.1108/RSR-07-2021-0043>
- Decker, E. N., & Chapman, K. (2022). Launching chat service during the pandemic: Inaugurating a new public service under emergency conditions. *Reference Services Review*, 50(2), 163-178. <https://doi.org/10.1108/RSR-08-2021-0051>
- George, L. E., & Birla, L. (2018). A study of topic modeling methods. In *2018 Second International Conference on Intelligent Computing and Control Systems* (pp. 109-113). IEEE. <https://doi.org/10.1109/ICCONS.2018.8663152>
- Hervieux, S. (2021). Is the library open? How the pandemic has changed the provision of virtual reference services. *Reference Services Review*, 49(3/4), 267-280. <https://doi.org/10.1108/RSR-04-2021-0014>

- Hutto, C. J. (2014). *VADER-Sentiment-Analysis* [Repository]. GitHub.
<https://github.com/cjhutto/vaderSentiment>
- Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216-225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- Kathuria, S. (2021). Library support in times of crisis: An analysis of chat transcripts during COVID-19. *Internet Reference Services Quarterly*, 25(3), 107-119. <https://doi.org/10.1080/10875301.2021.1960669>
- Koh, H., & Fienup, M. (2021). Topic modeling as a tool for analyzing library transcripts. *Information Technology and Libraries*, 40(3). <https://doi.org/10.6017/ital.v40i3.13333>
- Kohler, E. (2017). What do your library chats say?: How to analyze webchat transcripts for sentiment and topic extraction. In F. Baudino, K. Hart, & C. Johnson (Eds.), *17th Annual Brick & Click: An Academic Library Conference* (pp. 138-148). Northwest Missouri State University.
<https://eric.ed.gov/?id=ED578189>
- Kohler, E. (2020). Mining library chats: Sentiment analysis and topic extraction. In *14th International Conference on Performance Measurement in Libraries* (pp. 220-229).
- Lamba, M., & Madhusudhan, M. (2022). *Text mining for information professionals: An uncharted territory*. Springer. <https://doi.org/10.1007/978-3-030-85085-2>
- Liu, H., Chatterjee, I., Zhou, M., Lu, X. S., & Abusorrah, A. (2020). Aspect-based sentiment analysis: A survey of deep learning methods. *IEEE Transactions on Computational Social Systems*, 7(6), 1358-1375. <https://doi.org/10.1109/TCSS.2020.3033302>
- Lund, B. D. (2020). Assessing library topics using sentiment analysis in R: A discussion and code sample. *Public Services Quarterly*, 16(2), 112-123. <https://doi.org/10.1080/15228959.2020.1731402>
- Mawhinney, T. (2020). User preferences related to virtual reference services in an academic library. *The Journal of Academic Librarianship*, 46(1), 102094. <https://doi.org/10.1016/j.acalib.2019.102094>
- Munip, L., Tinik, L., Borrelli, S., Randone, G. R., & Paik, E. J. (2022). Lessons learned: A meta-synthesis examining library spaces, services and resources during COVID-19. *Library Management*, 43(1/2), 80-92. <https://doi.org/10.1108/LM-08-2021-0070>
- Murphy, J. E., Lewis, C. J., McKillop, C. A., & Stoeckle, M. (2022). Expanding digital academic library and archive services at the University of Calgary in response to the COVID-19 pandemic. *IFLA Journal*, 48(1), 83-98. <https://doi.org/10.1177/03400352211023067>
- Osorio, N., & Droog, A. (2021). Exploring the impact of the pandemic on reference and research services: A literature review. *New Review of Academic Librarianship*, 27(3), 280-300.
<https://doi.org/10.1080/13614533.2021.1990092>
- Ozeran, M., & Martin, P. (2019). “Good night, good day, good luck”: Applying topic modeling to chat reference transcripts. *Information Technology and Libraries*, 38(2), 49-57.
<https://doi.org/10.6017/ital.v38i2.10921>

- Paulus, T. M., Wise, A. F., & Singleton, R. (2019). How will the data be analyzed? Part one: Quantitative approaches including content analysis, statistical modeling, and computational methods. In T. M. Paulus & A. F. Wise (Eds.), *Looking for insight, transformation, and learning in online talk* (pp. 127-159). Routledge. <https://doi.org/10.4324/9781315283258>
- Radford, M. L., Costello, L., & Montague, K. (2020). Chat reference in the time of COVID-19: Transforming essential user services. *Association for Library and Information Science Education (ALISE) 2020 Conference*. <http://hdl.handle.net/2142/108820>
- Radford, M. L., Costello, L., & Montague, K. E. (2022). "Death of social encounters": Investigating COVID-19's initial impact on virtual reference services in academic libraries. *Journal of the Association for Information Science and Technology*, 73(11), 1594-1607. <https://doi.org/10.1002/asi.24698>
- Schiller, S. Z. (2016). CHAT for chat: Mediated learning in online chat virtual reference service. *Computers in Human Behavior*, 65, 651-665. <https://doi.org/10.1016/j.chb.2016.06.053>
- Sharma, A., Barrett, K., & Stapelfeldt, K. (2022). Natural language processing for virtual reference analysis. *Evidence Based Library and Information Practice*, 17(1), 78-93. <https://doi.org/10.18438/ebliip30014>
- Sobol, B., Goncalves, A., Vis-Dunbar, M., Lacey, S., Moist, S., Jantzi, L., Gupta, A., Mussell, J., Foster, P. L., & James, K. (2023). Chat transcripts in the context of the COVID-19 pandemic: Analysis of chats from the AskAway consortia. *Evidence Based Library and Information Practice*, 18(2), 73-92. <https://doi.org/10.18438/ebliip30291>
- Turp, C., & Hervieux, S. (2023). Exploring an automated method for the analysis of virtual reference interactions. *Reference Services Review*. Advance online publication. <https://doi.org/10.1108/RSR-05-2023-0050>
- Walker, J., & Coleman, J. (2021). Using machine learning to predict chat difficulty. *College & Research Libraries*, 82(5), 683-707. <https://doi.org/10.5860/crl.82.5.683>
- Wang, Y. (2022). Using machine learning and natural language processing to analyze library chat reference transcripts. *Information Technology and Libraries*, 41(3). <https://doi.org/10.6017/ital.v41i3.14967>
- Watson, A. P. (2023). Pandemic chat: A comparison of pandemic-era and pre-pandemic online chat questions at the University of Mississippi Libraries. *Internet Reference Services Quarterly*, 27(1), 25-36. <https://doi.org/10.1080/10875301.2022.2117757>
- Yatcilla, J. K., & Young, S. (2021). Library responses during the early days of the pandemic: A bibliometric study of the 2020 LIS literature. *Journal of Library Administration*, 61(8), 964-977. <https://doi.org/10.1080/0193082>