ARTICLE

Topic Modeling as a Tool for Analyzing Library Chat Transcripts

HyunSeung Koh and Mark Fienup

ABSTRACT

Library chat services are an increasingly important communication channel to connect patrons to library resources and services. Analysis of chat transcripts could provide librarians with insights into improving services. Unfortunately, chat transcripts consist of unstructured text data, making it impractical for librarians to go beyond simple quantitative analysis (e.g., chat duration, message count, word frequencies) with existing tools. As a stepping-stone toward a more sophisticated chat transcript analysis tool, this study investigated the application of different types of topic modeling techniques to analyze one academic library's chat reference data collected from April 10, 2015, to May 31, 2019, with the goal of extracting the most accurate and easily interpretable topics. In this study, topic accuracy and interpretability—the quality of topic outcomes—were quantitatively measured with topic coherence metrics. Additionally, qualitative accuracy and interpretability were measured by the librarian author of this paper depending on the subjective judgment on whether topics are aligned with frequently asked questions or easily inferable themes in academic library contexts. This study found that from a human's qualitative evaluation, Probabilistic Latent Semantic Analysis (pLSA) produced more accurate and interpretable topics, which is not necessarily aligned with the findings of the quantitative evaluation with all three types of topic coherence metrics. Interestingly, the commonly used technique Latent Dirichlet Allocation (LDA) did not necessarily perform better than pLSA. Also, semi-supervised techniques with human-curated anchor words of Correlation Explanation (CorEx) or guided LDA (GuidedLDA) did not necessarily perform better than an unsupervised technique of Dirichlet Multinomial Mixture (DMM). Last, the study found that using the entire transcript, including both sides of the interaction between the library patron and the librarian, performed better than using only the initial question asked by the library patron across different techniques in increasing the quality of topic outcomes.

INTRODUCTION

With the rise of online education, library chat services are an increasingly important tool for student learning.[1] Library chat services have the potential to support student learning, especially for distant learners who have a lack of opportunity to come and learn about library and research skills in person. In addition, unlike traditional in-person reference services whose use has declined drastically, library chat services have become an important communication channel that connects patrons to library resources, services, and spaces.[2]

Quantitative and qualitative analysis of chat transactions could provide librarians with insights into improving the quality of these resources, services, and spaces. For example, in order to maximize patrons' satisfaction, librarians could identify or evaluate quantitative and qualitative

HyunSeung Koh (hyunseung.koh@uni.edu) is an Assessment Librarian and Assistant Professor of Library Services, University of Northern Iowa. **Mark Fienup** (mark.fienup@uni.edu) is an Associate Professor in the Computer Science Department, University of Northern Iowa.

https://doi.org/10.6017/ital.v40i3.13333

patterns of chat reference data (e.g., busiest days and times of nondirectional, research-focused questions) and develop a better staffing plan for assigning librarians or student employees to most appropriate days and times. Furthermore, these insights could be used to help demonstrate library value by showing external stakeholders how successfully library chat services support students' needs, which is increasingly in demand for higher education.[3] In practice, it is burdensome for librarians to go beyond simple quantitative analysis (e.g., chat duration, message count, word frequencies) with existing chat software tools, such as LibraryH3lp, QuestPoint, Springshare's LibChat, and LivePerson.[4] Currently, in order to obtain rich and hidden insights from large volumes of chat transcripts, librarians need to conduct manual qualitative analysis of chat transcripts with unstructured text data, which requires a lot of time and effort.

In an age when library patrons' information needs have been changing, the lack of chat analysis tools that handle large volumes of transcripts hinders librarians' ability to respond to patrons' wants and needs in a timely manner.[5] In particular, small and medium-sized academic libraries have seen a shortage of librarians and need to hire and train student employees, so librarians' capabilities for real-time quick and easy analysis and assessment will become critical in helping them take appropriate actions to best meet user needs.[6] As part of an effort to develop a quick and easy analysis tool for large volumes of chat transcripts, this study applied topic modeling, which is a statistical technique "for learning the latent structure in document collections" or "a type of statistical model for finding hidden topical patterns of words."[7] We compared outcomes of different types of topic modeling techniques and attempted to propose topic modeling techniques that would be most appropriate in the context of chat reference transcript data.

LITERATURE REVIEW

To identify the most appropriate research methods that would facilitate analyzing a vast amount of chat transcripts, this section first introduces literature in relation to research methods used in analyzing chat transcript data in library settings and nonlibrary settings. It follows by discussing different types of topic modeling techniques that have high potential for quick and easy analysis of chat transcripts and their strengths and weaknesses.

Chat Transcript Analysis Methods in Library Settings

In analyzing library chat transcripts, which are one major data source of library chat service research, researchers have used variants of quantitative and qualitative research methods.[8] Coding-based content analysis with or without predefined categories is one type of qualitative method.[9] The other type of qualitative research method is conversation or language usage analysis but it is not a dominant type of research method, as compared to coding-based qualitative content analysis.[10] The most common quantitative methods are simple descriptive count- or frequency-based analyses that are accompanied by qualitative coding-based content analyses.[11] In some recent research, advanced quantitative research methods, such as cluster analysis and topic modeling techniques, have been used, but they have not been fully explored yet with a wide range of techniques.[12]

Chat Transcript Analysis Methods in Nonlibrary Settings

As shown in table 1, researchers in nonlibrary settings also used research methods in analyzing chat data from diverse technology platforms or contexts, ranging from qualitative manual coding methods to data mining and machine learning techniques. Topic modeling techniques are one of the chat analysis methods, but again, it seems that they have not been fully explored yet in chat analyses in nonlibrary settings, even though they have been used in a wide range of contexts.[13]

Table 1. Chat transcript analysis applications in non-library settings

Disciplines	Platforms/sources of chat transcript data	Chat transcript analysis methods/tools/techniques
Education	Chat rooms and text chat[14]	Qualitative content analysis
Health	Social media[15]	Qualitative & quantitative content analysis
Business	In-game chat features and chatbots[16]	A spell-checker, readability scores, the number of spelling and grammatical errors, Linguistic Inquiry and Word Count (LIWC) program, logistic regression analysis, Decision Tree, Support Vector Machine (SVM)
Criminology	Instant messengers, Internet Relay Chat (IRC) channels, internet-based chat logs, and social media[17]	LIWC program, cluster analysis, Latent Dirichlet Allocation (LDA)

Topic Modeling Techniques and Their Strengths and Weaknesses

As a quantitative and statistical method appropriate for analyzing a vast amount of chat transcript data, researchers from both library and nonlibrary settings used topic modeling. As shown in table 2, conventional topic modeling techniques include Latent Semantic Analysis, Probabilistic Latent Semantic Analysis, and Latent Dirichlet Allocation, each of which has its unique strengths and weaknesses.[18]

In order to overcome weaknesses of the conventional techniques, researchers have developed alternative techniques. For example, Dirichlet Multinomial Mixture (DMM) has been proposed to overcome data sparsity problems in short texts.[19] As another example, Correlation Explanation (CorEx) has been proposed to avoid time and effort to identify topics and their structure ahead of time.[20] Last, guided LDA (GuidedLDA) has been proposed to improve performance of infrequently occurring topics.[21]

Table 2. Strengths and weaknesses of conventional topic modeling techniques

	Acronym	Definitions	Strengths	Weaknesses
Latent Semantic Analysis	LSA	A document is represented as a vector of numbers found by applying dimensionality reduction (specifically, truncated SVD) to summarize the frequencies of co-occurring words across documents.	Can deal with polysemy (multiple meanings) to some extent.	Is hard to obtain and to determine the optimal number of topics.
Probabilistic Latent Semantic Analysis	pLSA	A document is represented as vectors, but these vectors have nonnegative entries summing to 1 such that each component (topic) represents the relative prominence of some probabilistic mixture of words in the corpus. Topics in a document are "probabilistic instead of the heuristic geometric distances."[22]	Can deal with polysemy issues; provides easy interpretation terms of word, document, and topic probabilities.	Has over-fitting problems.
Latent Dirichlet Allocation	LDA	A Bayesian extension of pLSA that adds assumptions about the relative probability of observing different document's distributions over topics.	Prevents over-fitting problems; provides a fully Bayesian probabilistic interpretation.	Does not show relationships among topics.

DATA, PREPROCESSING, ANALYSIS, AND EVALUATION

This section first introduces the data used for this study. Next, it explains the procedures of each stage starting from preprocessing to analyzing chat transcript data using different types of conventional and alternative topic modeling techniques. Last, it discusses quantitative and qualitative evaluation in terms of the quality of topic outcomes across different types of topic technique. For more details including Python scripts please visit our GitHub page at https://github.com/mfienup/uni-library-chat-study.

Data

This study collected the University of Northern Iowa's Rod Library chat reference data dated from April 10, 2015, to May 31, 2019 (IRB#18-0225). This raw chat data was downloaded from LibChat in the form of an Excel spreadsheet with 9,942 English chat transcripts with each transcript as a separate row.

Preprocessing

As the first step, this study removed unnecessary components of each chat transcript using a custom Python script. Components removed were timestamps, patron and librarian identifiers, http tags (e.g., URLs), and non-ASCII characters. Next, it processed the resulting text words using Python's Natural Language ToolKit (https://www.nltk.org/) and its WordNetLemmatizer function (https://www.nltk.org/_modules/nltk/stem/wordnet.html) to normalize words for further analyses. As the final step, it prepared the four types of data sets to identify which type of data set would produce better topic outcomes.

The four types of data sets were as follows:

- Question-Only: consists of only the initial question asked by the library patron in each chat transcript. Only the latter 10.7% of the chats recorded in the Excel spreadsheet contained an Initial Question column entry. The remaining chats assumed to contain their initial question in the patron's first response if it was longer than a trivial welcome message.
- Whole-Chat: consists of the whole chat transcripts from the library patron and librarians.
- Whole-Chat with Nouns and Adjectives: consists of only nouns and adjectives as parts of speech (POS) from the whole chat transcripts.
- Whole-Chat with Nouns, Adjectives, and Verbs: consists of only nouns, adjectives, and verbs as POS from the whole chat transcripts.

The first two data sets were prepared to examine if the first question initiated by each patron or the whole chat transcripts would help produce better topic outcomes. The last two data sets were prepared to examine which parts of speech retained would help produce better topic outcomes.

Data Analysis with Conventional Topic Modeling Techniques

This study first analyzed chat reference data using three conventional topic modeling techniques: Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (pLSA), and two versions of Latent Dirichlet Allocation (LDA), as shown in table 3.

All three techniques are examples of unsupervised topic modeling techniques that automatically analyze text data from a set of documents (in this study, a set of chat transcripts) to infer predominant topics or themes across all documents without human help.

A key challenge, or a key parameter to be determined, for unsupervised topic modeling techniques is to identify the optimal number of topics. The study ran the commonly used LDA technique with the Whole-Chat data set with various numbers of topics. Fifteen was chosen as an optimal number of topics for this study by calculating and comparing the log-likelihood scores among various number of topics.

Table 3. Conventional topic modeling techniques and their sources

Technique	Programming language	Implementation source	Version used in the study
Latent Semantic Analysis	Python	https://pypi.org/project/gensim/	3.8.1
Probabilistic Latent Semantic Analysis	Python	https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.NMF.html	0.21.3
Latent Dirichlet Allocation (with sklearn)	Python	https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.LatentDirichletAllocation.html	0.21.3
Latent Dirichlet Allocation (with PyMallet)	Python	https://github.com/mimno/PyMallet	Dated February 26, 2019

Also, before analyzing chat transcript data using LSA and pLSA, this study performed a term frequency–inverse document frequency (TF–IDF) transformation. TF–IDF is a measure of how important a word is to a document (i.e., a single chat transcript) compared to its relevance in a collection of all documents.

Data Analysis with Alternative Topic Modeling Techniques

In addition to conventional topic modeling techniques, this study analyzed chat reference data using three alternative techniques of Dirichlet Multinomial Mixture (DMM), anchored Correlation Explanation (CorEx) and guided LDA (GuidedLDA), as shown in table 4.

This study selected DMM as an alternative unsupervised topic modeling technique that has been developed for short texts. Also, this study selected anchored CorEx and guided LDA (GuidedLDA) as semi-supervised topic modeling techniques that require human-curated sets of words, called *anchors* or *seeds*, which nudge topic models toward including the suggested anchors. This is based on the assumption that human's curated techniques would help produce better quality of topics than the unsupervised techniques. For example, the three words "interlibrary," "loan," and "request," or the two words "article" and "database," are possible anchor words in the context of library chat transcripts. Such anchor words can appear anywhere within a chat in any order.

Table 4. Alternative topic modeling techniques and their sources

Unsupervised vs. semi-supervised	Technique	Programming language	Implementation source	Version used in the study
Unsupervised	Dirichlet Multinomial Mixture (DMM)	Java	https://github.com/qiang2100/STTM	9/27/2019
Semi-supervised	Anchored Correlation Explanation (CorEx)	Python	https://github.com/gregversteeg/corex_topic	1/21/2020
Semi-supervised	Guided LDA using collapsed Gibbs sampling	Python	https://guidedlda.readthedocs.io/en/latest/	10/5/2017

Given that a known set of anchor words associated with academic library chats seems unavailable in the literature, this study decided to obtain a list of most meaningful anchor words by combining outcomes of the unsupervised techniques with a human's follow-up curation, as follows:

Step 1. Execute unsupervised topic modeling techniques

Step 2. Combine resulting topics from all unsupervised topic modeling techniques

Step 3. Identify a list of all possible pairs of words (bi-occurrences), e.g., 28 pairs of words if each topic has 8 words, and all possible combinations of tri-occurrences of words

Step 4. Identify most common bi-occurrences and tri-occurrences of words across all topics by ordering in descending order by frequency

Step 5. Select a set of anchors from these bi-occurrences and tri-occurrences of words by a human's judgment

In terms of selecting a set of anchor words, the librarian author of this paper judged whether combinations of words in each row from step 4 were aligned with frequently asked questions or easily inferable themes in academic library contexts.

As shown in table 5, a set of "interlibrary," "loan," and "request" was selected as anchor words that are aligned with one frequently asked question about interlibrary loan requests, whereas a set of "access," "librarian," and "research" was not selected as anchor words because multiple themes, such as access to resources and asking for research help from librarians, can be inferred. Additionally, a set of "hour," "time," and "today" was selected over a set of "time," "tomorrow," and "tonight" as better or clearer anchor words.

Table 5. Examples of anchor words that were selected and not selected

	Examples of tri-occurrences of words (Note: Strikethrough denotes a set of words that were not selected as anchor words)		
1	interlibrary	loan	request
2	hour	time	today
~~3~~	~~time~~	~~tomorrow~~	~~tonight~~
~~4~~	~~time~~	~~today~~	~~tomorrow~~
5	floor	librarian	research
~~6~~	~~access~~	~~librarian~~	~~research~~
7	camera	digital	hub
~~8~~	~~digital~~	~~hub~~	~~medium~~
9	access	article	journal
10	access	article	database
11	access	account	campus
12	research	source	topic
~~13~~	~~paper~~	~~research~~	~~topic~~

Quantitative Evaluation with Topic Coherence Metrics

Comparing the quality of topic outcomes across various topic modeling techniques is tricky. Purely statistical and quantitative evaluation techniques, such as held-out log-likelihood measures, have proven to be unaligned with human intuition or judgment with respect to topic interpretability and coherency.[23] Thus, this study adopted the three topic coherence metrics of TC-PMI (normalized Pointwise Mutual Information), TC-LCP (normalized Log Conditional Probability), and TC-NZ (number of topic word pairs never observed together in the corpus) that have been introduced by Boyd-Graber, Mimno, and Newman; Bouma; and Lau, Newman, and Baldwin.[24] These three metrics are based on the assumption that the likelihood that two words that co-occur in a topic would also co-occur within a corpus.

To utilize the three topic coherence metrics, the study chose a binarized choice (e.g., Does a transcript contain two words?) instead of a sliding window of fixed size (e.g., Do two words appear within a fixed window of 10 consecutive words?) as a type of how to count term co-occurrences. This decision was made because each chat transcript is relatively short, and a fixed window size seemed inconsistent across different type of data sets that included different parts of speech.

In terms of the other decision to be made for applying the three topic coherence metrics, this study chose a training corpus of all the chat transcripts instead of external corpuses such as the entire collection of English Wikipedia articles that has little in common with average library chat transcripts.

Qualitative Evaluation with Human Judgment

In addition to quantitative evaluation with topic coherence metrics, qualitative accuracy and interpretability were judged by the librarian author of this paper based on whether topics were aligned with frequently asked questions or easily inferable themes in academic library contexts. For example, “Find or access book or article” was inferred, from a set of words in topic 1 on LSA in table 6, as an accurate and easily interpretable theme. From a set of words in topic 3 on LDA, “Reserve study room” and “Check out laptop computer” were inferred as two separable, easily interpretable themes. From a set of words in topic 15 on CorEx with nine anchors, no theme was inferred as an easily interpretable theme. (See table 10 in the Results section for all themes inferred from table 6.)

Table 6. Examples of topics found by topic modeling techniques

Topic modeling technique	Topics (Top 15 topics with eight words per topic) Note: Parenthetical additions are explanations or descriptions and not part of the topic.
Latent Semantic Analysis (LSA)	Topic 1. article book search find access link will check Topic 2. renew book article room reserve search journal check Topic 3. room renew reserve book study scheduler loan online Topic 4. renew request loan interlibrary search room review peer Topic 5. loan floor renew access interlibrary request log book Topic 6. book open print request search loan renew interlibrary Topic 7. print floor open printer color hour research pm Topic 8. open hour print search review close peer floor Topic 9. print access renew research book loan librarian open Topic 10. floor article open book renew print locate database Topic 11. article book attach file print database floor check Topic 12. check book desk laptop answer print shortly open Topic 13. answer desk shortly place room database circulation pick Topic 14. review peer search reserve log access campus database Topic 15. database file attach collection access journal research reserve
Probabilistic Latent Semantic Analysis (pLSA)	Topic 1. collection special youth contact email number archive department Topic 2. book title hold online check pick number reserve Topic 3. room reserve study scheduler reservation group rodscheduler (software) space Topic 4. search bar click type journal onesearch (a discovery tool) result homepage Topic 5. request loan interlibrary link illiad (system) submit inter instruction Topic 6. renew online account book today number circulation item Topic 7. access link log campus click work online sign Topic 8. article journal attach file title access google scholar Topic 9. research librarian paper appointment consultation source topic question Topic 10. open hour today close pm tomorrow midnight tonight Topic 11. check answer place shortly desk laptop student long Topic 12. print color printer computer printing mobile release black Topic 13. floor locate desk stack main fourth number section

Topic modeling technique	**Topics** **(Top 15 topics with eight words per topic)** Note: Parenthetical additions are explanations or descriptions and not part of the topic.
	Topic 14. database az subject ebsco(database) list business topic access Topic 15. review peer journal topic sociology study article result
Latent Dirichlet Allocation (LDA) with sklearn	Topic 1. file attach cite citation link article author pdf Topic 2. check book renew student item today time member Topic 3. room reserve computer laptop study check reservation desk Topic 4. book request loan interlibrary check title online copy Topic 5. search article database review result type google bar Topic 6. student class access iowa course university college fall Topic 7. research librarian source paper topic good appointment specific Topic 8. email contact chat good librarian work question address Topic 9. open hour today check pick hold desk close Topic 10. link access click log work campus sign database Topic 11. floor locate desk main art music circulation section Topic 12. medium digital check video hub desk rent camera Topic 13. article journal access title online link education amp Topic 14. print printer color card scan document charge job Topic 15. answer check place collection shortly special question number
Dirichlet Multinomial Mixture (DMM)	Topic 1. room reserve how will study check floor what Topic 2. request loan book interlibrary how article will link Topic 3. article access find journal link how search full Topic 4. book how find check what online link will Topic 5. article find attach file what how will link Topic 6. how check open today desk hour will what Topic 7. find article what search how research source database Topic 8. how print will cite printer link what citation Topic 9. search article find how review will database journal Topic 10. book find floor how will where call number Topic 11. book check how renew will today request what Topic 12. research how librarian find what article will email Topic 13. find how will contact collection what special email Topic 14. access article link log how campus database work Topic 15. article find will search what link book how
Anchored Correlation Explanation (CorEx) with nine anchor words	Topic 1. request loan interlibrary illiad (system) form submit inter fill Topic 2. study reserve room scheduler hub medium equipment digital

Topic modeling technique	**Topics** **(Top 15 topics with eight words per topic)** Note: Parenthetical additions are explanations or descriptions and not part of the topic.
	Topic 3. search review peer bar result type onesearch (a discovery tool) homepage Topic 4. today open hour pm assist close window midnight Topic 5. locate floor main where third fourth desk stack Topic 6. print printer color printing black white mobile release Topic 7. number collection special call phone youth archive xxx Topic 8. research librarian appointment consultation paper set xxx transfer Topic 9. access database journal article campus full az text Topic 10. email will contact work when good who student Topic 11. education read school class professor amp teacher child Topic 12. topic source cite write apa start citation recommend Topic 13. find attach file google what scholar title specific Topic 14. click log link left side catid button hand Topic 15. shortly place answer check cedar fall iowa northern
GuidedLDA with nine anchor words and confidence 0.75	Topic 1. book request loan interlibrary will how check link Topic 2. room reserve how check will desk study medium Topic 3. search article find how will database book review Topic 4. book check how renew today will hour open Topic 5. book floor find how check where call locate Topic 6. print how computer will printer color desk student Topic 7. contact collection will find email special how check Topic 8. research librarian find how what will email article Topic 9. article access link how log click database find Topic 10. article find how access what link attach file Topic 11. find chat copy how good online what will Topic 12. article find file attach what journal will work Topic 13. how check book answer place shortly what find Topic 14. book how find what sport link video textbook Topic 15. how cite what find citation author article source

RESULTS

This section first introduces which topic modeling techniques, as well as which type of data set, performed the best on each of the three topic coherence metrics. It follows by introducing which technique was the best according to human qualitative judgment.

Quantitative Evaluation with Topic Coherence Metrics

Given that for a topic coherence metric TC-PMI larger values mean more coherent topics, table 7 and its corresponding figure 1 show that CorEx with anchor words on the Whole-Chat performed best on TC-PMI. TF–IDF & pLSA on the Whole-Chat performed better than LDA on the Whole-Chat.

Given that for topic coherence metric TC-LCP larger values mean more coherent topics, table 8 and its corresponding figure 2 show that DMM on the Whole-Chat performed best on TC-LCP. TF–IDF & pLSA on the Whole-Chat performed better than LDA, even though LDA (PyMallet) on the Whole-Chat performed better than TC-IDF & pLSA on the Whole-Chat.

Given that for topic coherence metric TC-NZ smaller values mean more coherent topics, table 9 and its corresponding figure 3 show that TF–IDF & pLSA, LDA and LDA (PyMallet) on the Whole-Chat performed best on TC-NZ.

Table 7. TC-PMI comparison of topic modeling techniques on the four types of data sets (with top 15 topics with eight words per topic)

Topic modeling technique	**Whole-Chat**	**Whole-Chat (noun, adjective, verb)**	**Whole-Chat (noun, adjective)**	**Question-Only**
TF–IDF & LSA	-0.066	-0.061	-0.063	-0.429
TF–IDF & pLSA	0.508	0.321	0.494	-0.122
LDA (sklearn)	0.378	0.261	0.099	-0.995
LDA (PyMallet)	0.218	0.262	0.271	-0.091
DMM	0.136	0.22	0.285	0.109
CorEx without anchor words	0.47	0.497	0.396	-0.584
CorEx with nine anchor words	0.522	0.534	0.558	-0.401
GuidedLDA with nine anchor words and confidence 0.75	0.133	0.216	0.262	0.069

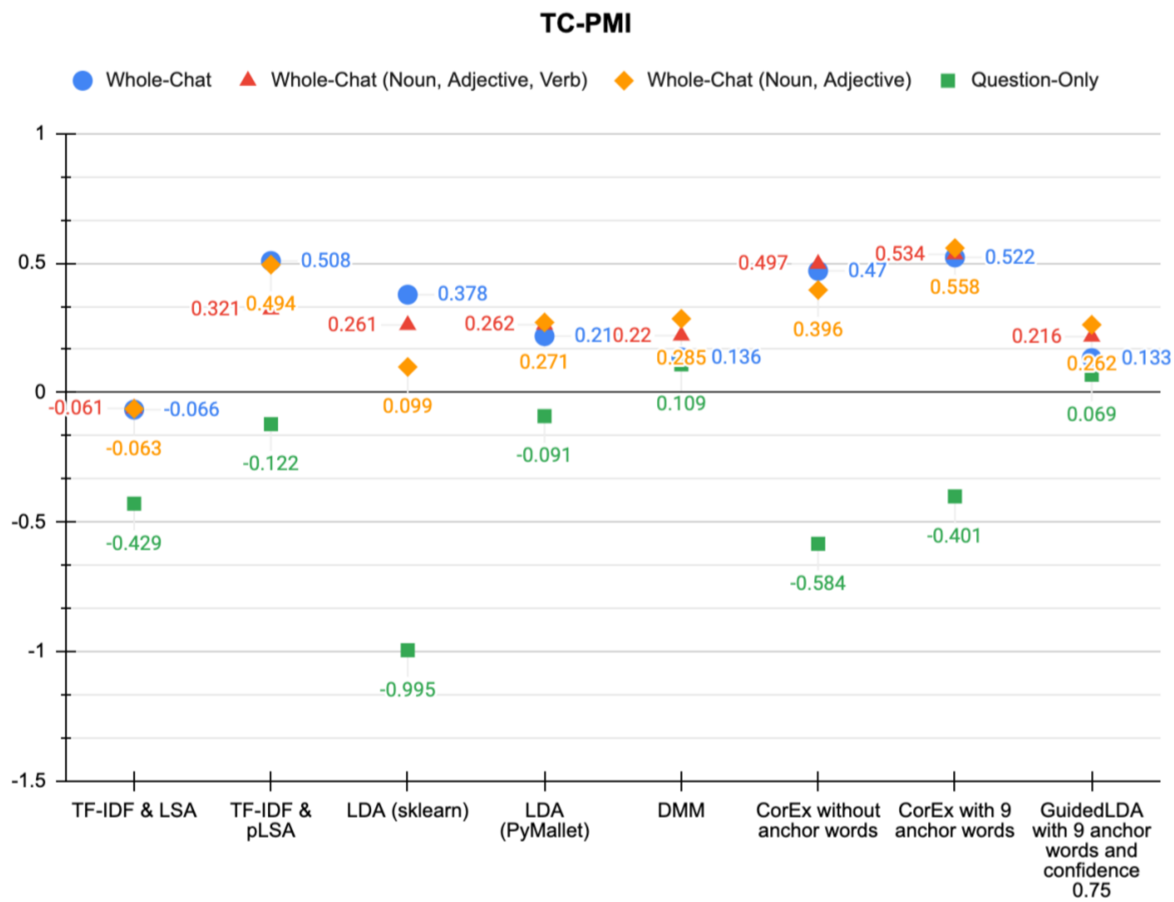

Figure 1. TC-PMI comparison of topic modeling techniques on the four types of data sets.

Table 8. TC-LCP comparison of topic modeling techniques on the four types of data sets (with top 15 topics with eight words per topic)

Topic modeling technique	Whole-Chat	Whole-Chat (noun, adjective, verb)	Whole-Chat (noun, adjective)	Question-Only
TF–IDF & LSA	-1.114	-1.124	-1.204	-1.675
TF–IDF & pLSA	-0.751	-0.793	-0.893	-1.956
LDA (sklearn)	-0.789	-0.979	-1.263	-2.827
LDA (PyMallet)	-0.637	-0.767	-0.918	-1.626
DMM	-0.546	-0.645	-0.731	-1.159
CorEx without anchor words	-0.868	-0.853	-1.062	-2.618
CorEx with nine anchor words	-0.82	-0.791	-0.884	-2.348
GuidedLDA with nine anchor words and confidence 0.75	-0.637	-0.686	-0.792	-1.143

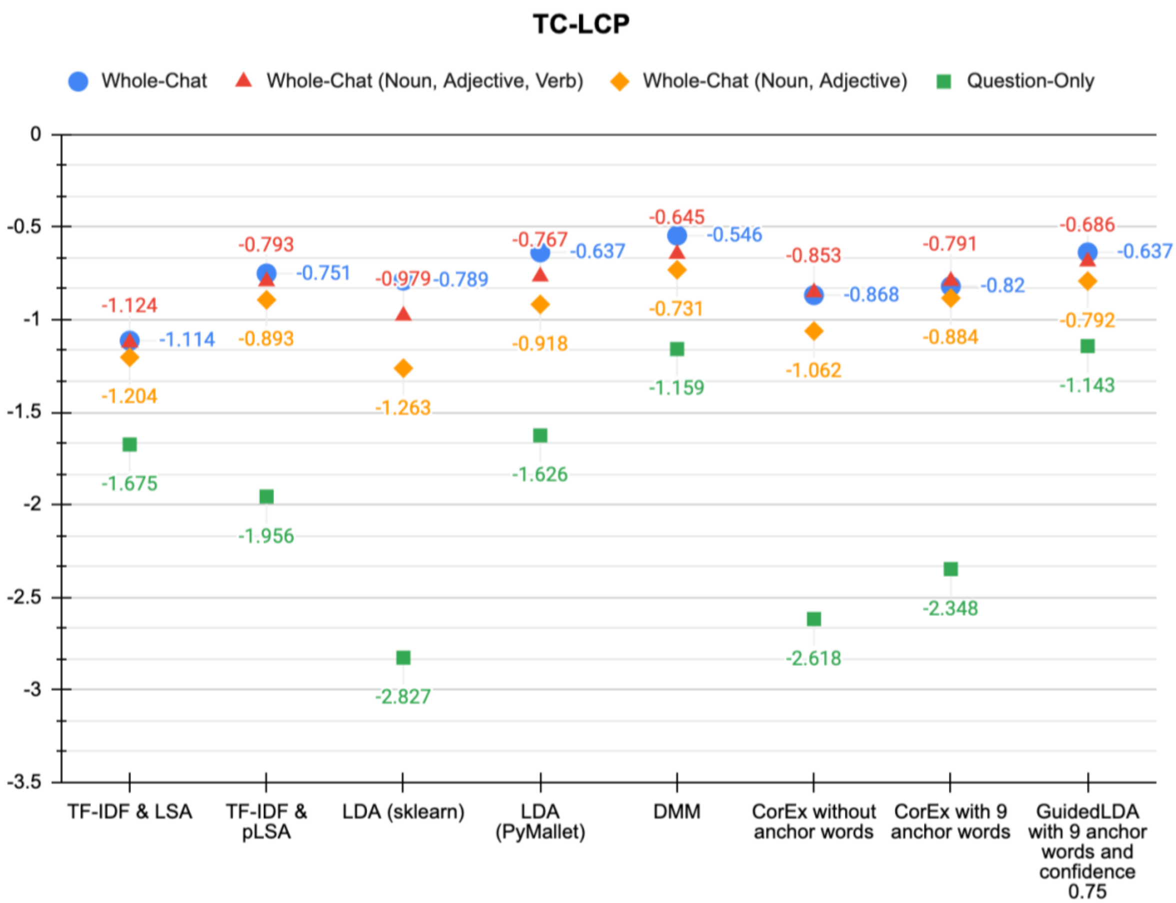

Figure 2. TC-LCP comparison of topic modeling techniques on the four types of data sets.

Table 9. TC-NZ comparison of topic modeling techniques on the four types of data sets (with top 15 topics with eight words per topic)

Topic modeling technique	Whole-Chat	Whole-Chat (noun, sdjective, verb)	Whole-Chat (noun, adjective)	Question-Only
TF–IDF & LSA	0.267	0.267	0.333	1.8
TF–IDF & pLSA	0	0	0.067	3.8
LDA (sklearn)	0	0.467	1.2	7.067
LDA (PyMallet)	0	0.133	0.267	1.8
DMM	0.067	0	0	0.267
CorEx without anchor words	0.333	0.067	0.6	7.067
CorEx with nine anchor words	0.133	0	0.133	5.267
GuidedLDA with nine anchor words and confidence 0.75	0.2	0.067	0	0.133

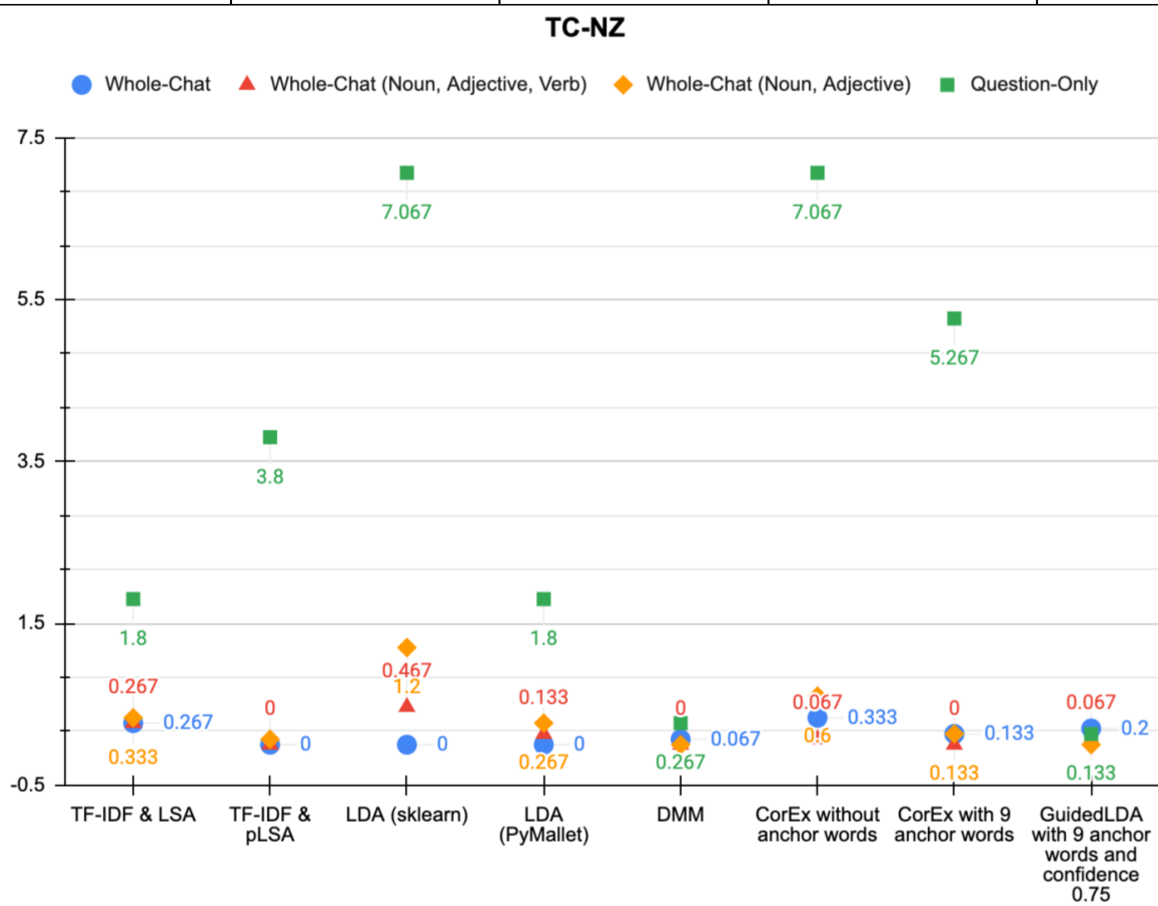

Figure 3. TC-NZ comparison of topic modeling techniques on all four data sets.

Last, all tables 7 to 9 and their corresponding figures 1 to 3 clearly show that the Whole-Chat data set with all parts of speech was generally the best data set on all the techniques.

Qualitative Evaluation with Human Judgment

As shown in table 10, all techniques had relatively high accuracy and interpretability in terms of straightforward topics or themes in italicized text, such as “interlibrary loan,” “technology,” “hours,” and “room reservations,” where one keyword could represent a whole theme. However, in terms of less-straightforward topics or themes pLSA performed better than the other techniques. In other words, pLSA had the highest number of topics that are aligned clearly with frequently asked questions or are easily inferable themes in academic library contexts. Also, pLSA had a lower number of unrelated or multiple themes within one topic, whereas other techniques had a higher number of unrelated or multiple themes within one topic. As an example, topic 8 on DMM shows that “print” and “citation” can be inferred as two unrelated themes within one topic.

Table 10. Examples of themes qualitatively inferred from a list of words (a topic) identified by each topic modeling technique

Topic modeling technique	**Themes inferred from table 6** (Note: Italics denotes straightforward themes; and strikethrough denotes themes with no interpretability or unrelated, multiple themes within one topic)
Latent Semantic Analysis (LSA)	Topic 1. Find or access book or article ~~Topic 2. Renew book or article; reserve a room; search journal~~ ~~Topic 3. Renew book online; reserve room; loan~~ ~~Topic 4. Renew; interlibrary loan; search; room~~ ~~Topic 5. Renew book; interlibrary loan; floor~~ ~~Topic 6. Renew; interlibrary loan print book; search~~ ~~Topic 7. Print color; floor; hours; research~~ ~~Topic 8. Hours; print; search; peer peer review; floor~~ ~~Topic 9. Print; renew book; librarian; open hours~~ ~~Topic 10. Renew book and article, print, floor and locate; database~~ ~~Topic 11. Print; database; floor~~ ~~Topic 12. Check out book or laptop; print; open~~ ~~Topic 13. Circulation desk; room; database~~ ~~Topic 14. Not clear~~ ~~Topic 15. Not clear~~
Probabilistic Latent Semantic Analysis (pLSA)	Topic 1. Contact information of special collection and Youth ~~Topic 2. Not clear~~ *Topic 3. Room reservation* Topic 4. Journal search and OneSearch *Topic 5. Interlibrary loan request* Topic 6. How to renew book online ~~Topic 7. Working from off campus (not clear)~~ Topic 8. Journal article via Google Scholar Topic 9. Appointment with librarians for research consultations *Topic 10. Open hours* ~~Topic 11. Not clear~~ *Topic 12. Printing*

Topic modeling technique	**Themes inferred from table 6** (Note: Italics denotes straightforward themes; and strikethrough denotes themes with no interpretability or unrelated, multiple themes within one topic)
	Topic 13. Stack on the fourth floor Topic 14. Databases A-Z for business including EBSCO Topic 15. Peer reviewed journals for Sociology
Latent Dirichlet Allocation (LDA) with sklearn	~~Topic 1. Not clear~~ ~~Topic 2. Not clear~~ ~~Topic 3. Reserve study room; check out laptop computer~~ *Topic 4. Interlibrary loan online* Topic 5. Search article via databases ~~Topic 6. Not clear~~ Topic 7. Appointment with research librarians Topic 8. Contact librarian via email *Topic 9. Open hours* Topic 10. Database access from off campus Topic 11. Floor for art and music circulation desk *Topic 12. Rent camera* Topic 13. Access journal article *Topic 14. Printing and charge* Topic 15. Special collection
Dirichlet Multinomial Mixture (DMM)	*Topic 1. Reserve study room and floor* *Topic 2. Interlibrary loan* Topic 3. Search and access article Topic 4. Find book online ~~Topic 5. Find article (Not clear)~~ *Topic 6. Open hours* Topic 7. Find article and database ~~Topic 8. Print; citation~~ Topic 9. Find article & database Topic 10. Find book with call number ~~Topic 11. Renew book (Not clear)~~ Topic 12. Email librarians for research help ~~Topic 13. Special collection (Not clear)~~ Topic 14. Access article/database from on campus ~~Topic 15. Find article (Not clear)~~
Anchored Correlation Explanation (CorEx) with nine anchor words	*Topic 1. Interlibrary loan* ~~Topic 2. Reserve study room; equipment~~ Topic 3. Peer-reviwed and OneSearch *Topic 4. Open hours* *Topic 5. Floor location* *Topic 6. Printing* Topic 7. Special collection and phone number Topic 8. Research consultation appointment Topic 9. Access database A-Z

Topic modeling technique	**Themes inferred from table 6** (Note: Italics denotes straightforward themes; and strikethrough denotes themes with no interpretability or unrelated, multiple themes within one topic)
	~~Topic 10. Not clear~~ ~~Topic 11. Not clear~~ Topic 12. APA citations ~~Topic 13. Google Scholar (Not clear)~~ Topic 14. Log in ~~Topic 15. Not clear~~
GuidedLDA with nine anchor words and confidence 0.75	*Topic 1. Interlibrary loan* *Topic 2. Reserve study room & medium* ~~Topic 3. Search and find article; databases~~ ~~Topic 4. Renew book; hours~~ Topic 5. Find book with call number *Topic 6. Printing* Topic 7. Special collection Topic 8. Email to research librarian Topic 9. Access article and databases ~~Topic 10. Access article; attach file (Not clear)~~ ~~Topic 11. Not clear~~ ~~Topic 12. Find article and journal; file attach (Not clear)~~ ~~Topic 13. Not clear~~ Topic 14. Find book, video, and textbook about sport Topic 15. Citation

DISCUSSION

Given that different topic modeling techniques performed the best depending on different types of topic coherence metrics, it is not possible to make a firm conclusion that one technique is better than the others. Interestingly, the commonly-used technique LDA tested in both sklearn and PyMallet in this study did not consistently outperform TF–IDF & pLSA. In addition, semi-supervised techniques of anchored Correlation Explanation (CorEx) or guided LDA (GuidedLDA) did not necessarily outperform an unsupervised technique of the Dirichlet Multinomial Mixture (DMM). Last, from a human's qualitative judgment, pLSA performed the best, which is aligned with the findings on TC-NZ. This might imply that TC-NZ is a more appropriate metric than the other metrics in measuring topic coherence in the context of academic library chat transcripts.

In terms of different types of data sets, all three of the Whole-Chat data sets significantly outperformed the Questions-Only data set. At the outset of the study, it was conjectured that the initial question of each chat transaction might concentrate the essence of each chat, thereby leading to better performance. Clearly this was not the case, possibly because the rest of chat transcripts would reinforce a topic by standardizing the vocabulary of the chat's initial question. It was somewhat interesting that varying the parts of speech (POS) retained in the three Whole-Chat data sets had little benefit on the topic modeling analyses. It might imply that topic modeling techniques are sensitive enough to differentiate across different parts of speech, thereby leading to good performance regardless of types of data sets.

CONCLUSION

This study clearly showed that conventional techniques should be also examined to avoid any errors from the assumption that newly developed techniques such as LDA would always outperform regardless of contexts. Also, both quantitative and qualitative evaluations indicate that unsupervised techniques should be equally weighted as semi-supervised techniques with human interventions. As a future study, like other similar research, it would be meaningful to compare human qualitative judgment with scores of each metric more rigorously, along with more librarians' input, to confirm (or disconfirm) our preliminary conclusion that TC-NZ is the most appropriate topic coherence metric in the context of library chat transcripts.[25] It would be also interesting to investigate and examine semi-supervised techniques with different types of anchoring approaches, such as tandem anchoring.[26] Last, in order to overcome limitations of this study, it would be valuable to collect more and diverse chat reference data and compare output of topics across different types of institutions (e.g., teaching versus research institutions).

ACKNOWLEDGMENTS

This project was made possible in part by the Institute of Museum and Library Services [National Leadership Grants for Libraries, LG-34-19-0074-19].

ENDNOTES

[1] Christina M. Desai and Stephanie J. Graves, "Cyberspace or Face-to-face: The Teachable Moment and Changing Reference Mediums," *Reference & User Services Quarterly* 47, no. 3 (Spring 2008): 242–55, https://www.jstor.org/stable/20864890; Megan Oakleaf and Amy Vanscoy, "Instructional Strategies for Digital Reference: Methods to Facilitate Student Learning," *Reference & User Services Quarterly* 49, no. 4 (Summer 2010): 380–90, https://www.jstor.org/stable/20865299; Shu Z. Schiller, "Chat for Chat: Mediated Learning in Online Chat Virtual Reference Service," *Computers in Human Behavior* 65 (July 2016): 651–65, https://doi.org/10.1016/j.chb.2016.06.053; Mila Semeshkina, "Five Major Trends in Online Education to Watch out for in 2021," *Forbes*, February 2, 2021, https://www.forbes.com/sites/forbesbusinesscouncil/2021/02/02/five-major-trends-in-online-education-to-watch-out-for-in-2021/?sh=3261272521eb.

[2] Maryvon Côté, Svetlana Kochkina, and Tara Mawhinney, "Do You Want to Chat? Reevaluating Organization of Virtual Reference Service at an Academic Library," *Reference and User Services Quarterly* 56, no. 1 (Fall 2016): 36–46, https://www.jstor.org/stable/90009882; Sarah Lemire, Lorelei Rutledge, and Amy Brunvand, "Taking a Fresh Look: Reviewing and Classifying Reference Statistics for Data-driven Decision Making," *Reference & User Services Quarterly* 55, no. 3 (Spring 2016): 230–38, https://www.jstor.org/stable/refuseserq.55.3.230; B. Jane Scales, Lipi Turner-Rahman, and Feng Hao, "A Holistic Look at Reference Statistics: Whither Librarians?," *Evidence Based Library and Information Practice* 10, no. 4 (December 2015): 173–85, https://doi.org/10.18438/B8X01H.

[3] Pamela J. Howard, "Can Academic Library Instant Message Transcripts Provide Documentation of Undergraduate Student Success?," *Journal of Web Librarianship* 13, no. 1 (February 2019): 61–87, https://doi.org/10.1080/19322909.2018.1555504.

[4] Côté and Kochkina, "Do You Want to Chat?"; Sharon Q. Yang and Heather A. Dalal, "Delivering Virtual Reference Services on the Web: An Investigation into the Current Practice by Academic Libraries," *Journal of Academic Librarianship* 41, no. 1 (November 2015): 68–86, https://doi.org/10.1016/j.acalib.2014.10.003.

[5] Feifei Liu, "How Information-Seeking Behavior Has Changed in 22 Years," NN/g Nielsen Norman Group, January 26, 2020, https://www.nngroup.com/articles/information-seeking-behavior-changes/; Amanda Spink and Jannica Heinström, eds., *New Directions in Information Behavior* (Bingley, UK: Emerald Group Publishing Limited, 2011).

[6] Kathryn Barrett and Amy Greenberg, "Student-Staffed Virtual Reference Services: How to Meet the Training Challenge," *Journal of Library & Information Services in Distance Learning* 12, no. 3–4 (August 2018): 101–229, https://doi.org/10.1080/1533290X.2018.1498620; Robin Canuel et al., "Developing and Assessing a Graduate Student Reference Service," *Reference Services Review* 47, no. 4 (November 2019): 527–43, https://doi.org/10.1108/RSR-06-2019-0041.

[7] Bhagyashree Vyankatrao Barde and Anant Madhavrao Bainwad, "An Overview of Topic Modeling Methods and Tools," in *Proceedings of International Conference on Intelligent Computing and Control Systems,* 2018, 745–50, https://doi.org/10.1109/ICCONS.2017.8250563; Jordan Boyd-Graber, David Mimno, and David Newman, "Care and Feeding of Topic Models: Problems, Diagnostics, and Improvements," in *Handbook of Mixed Membership Models and Their Application*s, eds. Edoardo M. Airoldi et al. (New York: CRC Press, 2014), 225–54.

[8] Miriam L. Matteson, Jennifer Salamon, and Lindy Brewster, "A Systematic Review of Research on Live Chat Service," *Reference & User Services Quarterly* 51, no. 2 (Winter 2011): 172–89, https://www.jstor.org/stable/refuseserq.51.2.172.

[9] Kate Fuller and Nancy H. Dryden, "Chat Reference Analysis to Determine Accuracy and Staffing Needs at One Academic Library," *Internet Reference Services Quarterly* 20, no. 3–4 (December 2015): 163–81, https://doi.org/10.1080/10875301.2015.1106999; Sarah Passonneau and Dan Coffey, "The Role of Synchronous Virtual Reference in Teaching and Learning: A Grounded Theory Analysis of Instant Messaging Transcripts," *College & Research Libraries* 72, no. 3 (2011): 276–95, https://doi.org/10.5860/crl-102rl.

[10] Paula R. Dempsey, "'Are You A Computer?' Opening Exchanges in Virtual Reference Shape the Potential for Teaching," *College & Research Librarie*s 77, no. 4 (2016): 455–68, https://doi.org/10.5860/crl.77.4.455; Jennifer Waugh, "Formality in Chat Reference: Perceptions of 17- to 25-year-old University Students," *Evidence Based Library and Information Practice* 8, no. 1 (2013): 19–34, https://doi.org/10.18438/B8WS48.

[11] Robin Brown, "Lifting the Veil: Analyzing Collaborative Virtual Reference Transcripts to Demonstrate Value and Make Recommendations for Practice," *Reference & User Services Quarterly* 57, no. 1 (Fall 2017): 42–47, https://www.jstor.org/stable/90014866; Sarah

Maximiek, Elizabeth Brown, and Erin Rushton, "Coding into the Great Unknown: Analyzing Instant Messaging Session Transcripts to Identify User Behaviors and Measure Quality of Service," *College & Research Libraries* 71, no. 4 (2010): 361–73, https://doi.org/10.5860/crl-48r1.

[12] Christopher Brousseau, Justin Johnson, and Curtis Thacker, "Machine Learning Based Chat Analysis," *Code4Lib Journal* 50 (February 2021), https://journal.code4lib.org/articles/15660; Ellie Kohler, "What Do Your Library Chats Say?: How to Analyze Webchat Transcripts for Sentiment and Topic Extraction," in *Brick & Click Libraries Conference Proceedings* (Maryville, MO: Northwest Missouri State University, 2017), 138–48, https://files.eric.ed.gov/fulltext/ED578189.pdf; Megan Ozeran and Piper Martin, "'Good Night, Good Day, Good Luck,'" *Information Technology and Libraries* 38, no. 2 (June 2019): 49–57, https://doi.org/10.6017/ital.v38i2.10921; Thomas Stieve and Niamh Wallace, "Chatting While You Work: Understanding Chat Reference User Needs Based on Chat Reference Origin," *Reference Services Review* 46, no. 4 (November 2018): 587–99, https://doi.org/10.1108/RSR-09-2017-0033; Nadaleen Tempelman-Kluit and Alexa Pearce, "Invoking the User from Data to Design," *College & Research Libraries* 75, no. 5 (2014): 616–40, https://doi.org/10.5860/crl.75.5.616.

[13] Jordan Boyd-Graber, Yuening Hu, and David Mimno, "Applications of Topic Models," *Foundations and Trends in Information Retrieval* 11, no. 2–3 (2017): 143–296, https://mimno.infosci.cornell.edu/papers/2017_fntir_tm_applications.pdf.

[14] Ewa M. Golonka, Medha Tare, and Carrie Bonilla, "Peer Interaction in Text Chat: Qualitative Analysis of Chat Transcripts," *Language Learning & Technology* 21, no. 2 (June 2017): 157–78, http://hdl.handle.net/10125/44616; Laura D. Kassner and Kate M. Cassada, "Chat It Up: Backchanneling to Promote Reflective Practice among In-Service Teachers," *Journal of Digital Learning in Teacher Education* 33, no. 4 (August 2017): 160–68, https://doi.org/10.1080/21532974.2017.1357512.

[15] Eradah O. Hamad et al., "Toward a Mixed-methods Research Approach to Content Analysis in the Digital Age: The Combined Content-analysis Model and Its Applications to Health Care Twitter Feeds," *Journal of Medical Internet Research* 18, no. 3 (March 2016): e60, https://doi.org/10.2196/jmir.5391; Janet Richardson et al., "Tweet If You Want to Be Sustainable: A Thematic Analysis of a Twitter Chat to Discuss Sustainability in Nurse Education," *Journal of Advanced Nursing* 72, no. 5 (January 2016): 1086–96, https://doi.org/10.1111/jan.12900.

[16] Shuyuan Mary Ho et al., "Computer-mediated Deception: Strategies Revealed by Language-Action Cues in Spontaneous Communication," *Journal of Management Information Systems* 33, no. 2 (October 2016): 393–420, https://doi.org/10.1080/07421222.2016.1205924; Mina Park, Milam Aiken, and Laura Salvador, "How Do Humans Interact with Chatbots?: An Analysis of Transcripts," *International Journal of Management & Information Technology* 14 (2018): 3338–50, https://doi.org/10.24297/ijmit.v14i0.7921.

[17] Abdur Rahman, M. A. Basher, and Benjamin C. M. Fung, "Analyzing Topics and Authors in Chat Logs for Crime Investigation," *Knowledge and Information Systems* 39, no. 2 (March 2014): 351–81, https://doi.org/10.1007/s10115-013-0617-y; Michelle Drouin et al., "Linguistic

Analysis of Chat Transcripts From Child Predator Undercover Sex Stings," *Journal of Forensic Psychiatry & Psychology* 28, no. 4 (February 2017): 437–57, https://doi.org/10.1080/14789949.2017.1291707; Da Kuang, P. Jeffrey Brantingham, and Andrea L. Bertozzi, "Crime Topic Modeling," *Crime Science* 6, no. 12 (December 2017): 1–12, https://doi.org/10.1186/s40163-017-0074-0; Md Waliur Rahman Miah, John Yearwood, and Siddhivinayak Kulkarni, "Constructing an Inter-post Similarity Measure to Differentiate the Psychological Stages in Offensive Chats," *Journal of the Association for Information Science and Technology* 66, no. 5 (January 2015): 1065–81, https://doi.org/10.1002/asi.23247.

[18] Charu C. Aggarwal and ChengXiang Zhai, eds. *Mining Text Data* (New York: Springer, 2012); Rubayyi Alghamdi and Khalid Alfalqi, "A Survey of Topic Modeling in Text Mining," *International Journal of Advanced Computer Science and Applications* 6, no. 1 (2015): 146–53, https://doi.org/10.14569/IJACSA.2015.060121; Leticia H. Anaya, "*Comparing Latent Dirichlet Allocation and Latent Semantic Analysis as Classifiers*" (PhD diss., University of North Texas, 2011); Barde and Bainwad, "An Overview of Topic Modeling"; David M. Blei, "Topic Modeling and Digital Humanities," *Journal of Digital Humanities* 2, no. 1 (Winter 2012), http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/; Tse-hsun Chen, Stephen W. Thomas, and Ahmed E. Hassan, "A Survey on the Use of Topic Models When Mining Software Repositories," *Empirical Software Engineering* 21, no. 5 (September 2016): 1843–919, https://doi.org/10.1007/s10664-015-9402-8; Elisabeth Günther and Thorsten Quandt, "Word Counts and Topic Models: Automated Text Analysis Methods for Digital Journalism Research," *Digital Journalism* 4, no. 1 (October 2016): 75–88, https://doi.org/10.1080/21670811.2015.1093270; Gabe Ignatow and Rada Mihalcea, *An Introduction to Text Mining: Research Design, Data Collection, and Analysis* (New York: Sage, 2017); Stefan Jansen, *Hands-on Machine Learning for Algorithmic Trading: Design and Implement Investment Strategies based on Smart Algorithms that Learn from Data Using Python* (Birmingham: Packt Publishing Limited, 2018); Lin Liu et al., "An Overview of Topic Modeling and Its Current Applications in Bioinformatics," *Springerplus* 5, no. 1608 (September 2016): 1–22, https://doi.org/10.1186/s40064-016-3252-8; John W. Mohr and Petko Bogdanov, "Introduction—Topic Models: What They Are and Why They Matter," *Poetics* 41, no. 6 (December 2013): 545–69, https://doi.org/10.1016/j.poetic.2013.10.001; Gerard Salton, Anita Wong, and Chung-shu Yang, "A Vector Space Model for Automatic Indexing," *Communications of the ACM* 18, no. 11 (November 1975): 613–20, https://doi.org/10.1145/361219.361220; Jianhua Yin and Jianyong Wang, "A Dirichlet Multinomial Mixture Model-based Approach for Short Text Clustering," in *Proceedings of the Twentieth ACM SIGKDD International Conference On Knowledge Discovery and Data Mining* (New York: ACM, 2014), 233–42, https://doi.org/10.1145/2623330.2623715; Hongjiao Xu et al., "Exploring Similarity between Academic Paper and Patent Based on Latent Semantic Analysis and Vector Space Model," in *Proceedings of the Twelfth International Conference on Fuzzy Systems and Knowledge Discovery* (New York: IEEE, 2015), 801–5, https://doi.org/10.1109/FSKD.2015.7382045; Chengxiang Zhai, *Statistical Language Models for Information Retrieval* (Williston, VT: Morgan & Claypool Publishers, 2018).

[19] Neha Agarwal, Geeta Sikkaa, and Lalit Kumar Awasthib, "Evaluation of Web Service Clustering Using Dirichlet Multinomial Mixture Model Based Approach for Dimensionality Reduction in Service Representation," *Information Processing & Management* 57, no. 4 (July 2020), https://doi.org/10.1016/j.ipm.2020.102238 ; Chenliang Li et al., "Topic Modeling for Short

Texts With Auxiliary Word Embeddings," in *Proceedings of the Thirty-Ninth International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York: ACM, 2016), 165–74, https://doi.org/10.1145/2911451.2911499; Jipeng Qiang et al., "Short Text Topic Modeling Techniques, Applications, and Performance: A Survey," *IEEE Transactions on Knowledge and Data Engineering* 14, no. 8 (April 2019): 1–17, https://doi.org/10.1109/TKDE.2020.2992485.

[20] Ryan J. Gallagher et al., "Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge," *Transactions of the Association for Computational Linguistics* 5 (December 2017): 529–42, https://doi.org/10.1162/tacl_a_00078.

[21] Jagadeesh Jagarlamudi, Hal Daumé III, and Raghavendra Udupa, "Incorporating Lexical Priors into Topic Models," in *Proceedings of the Thirteenth Conference of the European Chapter of the Association for Computational Linguistics* (Stroudsburg, PA: ACL, 2012), 204–13, https://www.aclweb.org/anthology/E12-1021; Olivier Toubia et al., "Extracting Features of Entertainment Products: A Guided Latent Dirichlet Allocation Approach Informed by the Psychology of Media Consumption," *Journal of Marketing Research* 56, no. 1 (December 2019): 18–36, https://doi.org/10.1177/0022243718820559.

[22] Nan Zhang and Baojun Ma, "Constructing a Methodology toward Policy Analysts for Understanding Online Public Opinions: A Probabilistic Topic Modeling Approach," in *Electronic Government and Electronic Participation*, eds. Efthimios Tambouris et al. (Amsterdam, Netherlands: IOS Press BV, 2015): 72–9, https://doi.org/10.3233/978-1-61499-570-8-72.

[23] Jonathan Chang et al., "Reading Tea Leaves: How Humans Interpret Topic Models," in *Proceedings of the Twenty-Second International Conference on Neural Information Processing Systems* (New York: ACM, 2009), 288–96, https://dl.acm.org/doi/10.5555/2984093.2984126.

[24] Gerlof Bouma, "Normalized (Pointwise) Mutual Information in Collocation Extraction," in *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology* (Tübingen, Germany: Gunter Narr Verlag, 2009), 43–53; Boyd-Graber, Mimno, and Newman, "Care and Feeding of Topic Models," in *Handbook of Mixed Membership Models and Their Applications*, eds. Edoardo M. Airoldi, David M. Blei, Elena A. Erosheva, and Stephen E. Fienberg (Boca Raton: CRC Press, 2014), 225–54; Jey Han Lau, David Newman, and Timothy Baldwin, "Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality," in *Proceedings of the Fourteenth Conference of the European Chapter of the Association for Computational Linguistics* (Stroudsburg, PA: ACL, 2014), 530–39, https://doi.org/10.3115/v1/E14-1056.

[25] Lau, Newman, and Baldwin, "Machine Reading Tea Leaves"; David Newman et al., "Automatic Evaluation of Topic Coherence," in *Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (New York: ACM, 2010), 100–108, https://dl.acm.org/doi/10.5555/1857999.1858011.

[26] Jeffrey Lund et al., "Tandem Anchoring: a Multiword Anchor Approach for Interactive Topic Modeling," in *Proceedings of the Fifty-fifth Annual Meeting of the Association for Computational Linguistics* (Stroudsburg, PA: ACL, 2017), 896–905, https://doi.org/10.18653/v1/P17-1083.