TEXT ANALYSIS ASSESSMENT

Using quantitative text analysis to assess and improve chat reference service

Alexander.Justice@lmu.edu

Loyola Marymount University, Los Angeles

THANK YOU

This research was a department and team effort, and could not have been carried out without a lot of assistance and advice from many quarters. At LMU: Susan Archambault, Javier Garibay, Melanie Hubbard, Burney Wong.

Even more thanks to Kathryn Ryan.

Special thanks to:

Scott Kleinman
Director, Center for the Digital Humanities
California State University, Northridge

Scott Enderle
Digital Humanities Specialist Librarian
University of Pennsylvania Libraries

Pam Sessoms & Amy Shelton, Nub Games, Inc. / libraryh3lp

CHAT REFERENCE

- Part of LMU library reference service since 2004

- We provide chat reference 24/7 by working with a consortium of most other Jesuit universities in US

- Our platform is libraryh3lp, a product of Nub Games, Inc.

- Libraryh3lp is set up to save transcripts of all chat sessions

THE DATA

- We converted six semesters' worth of chat transcripts into a lexical corpus for a variety of text analysis research.

- January 2014 to December 2016

- Corpus represents over 10K interactions

- Two sub-corpora: librarians and guests (99% students)
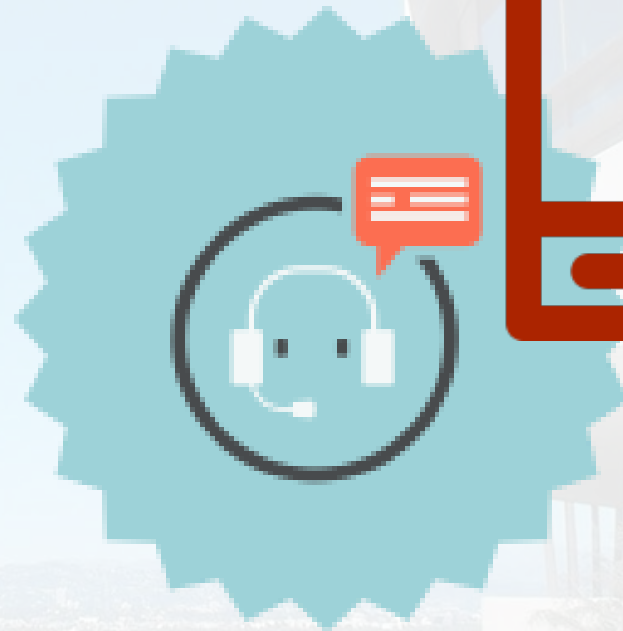
THE RESEARCH QUESTIONS

- R1: What terminology and vocabulary do students use to describe research tools and resources (compared to librarian terminology)?

- R2: What feelings or sentiments are conveyed by students during chat?

- R3: What are the frequent question parameters and categories, and what drives traffic?
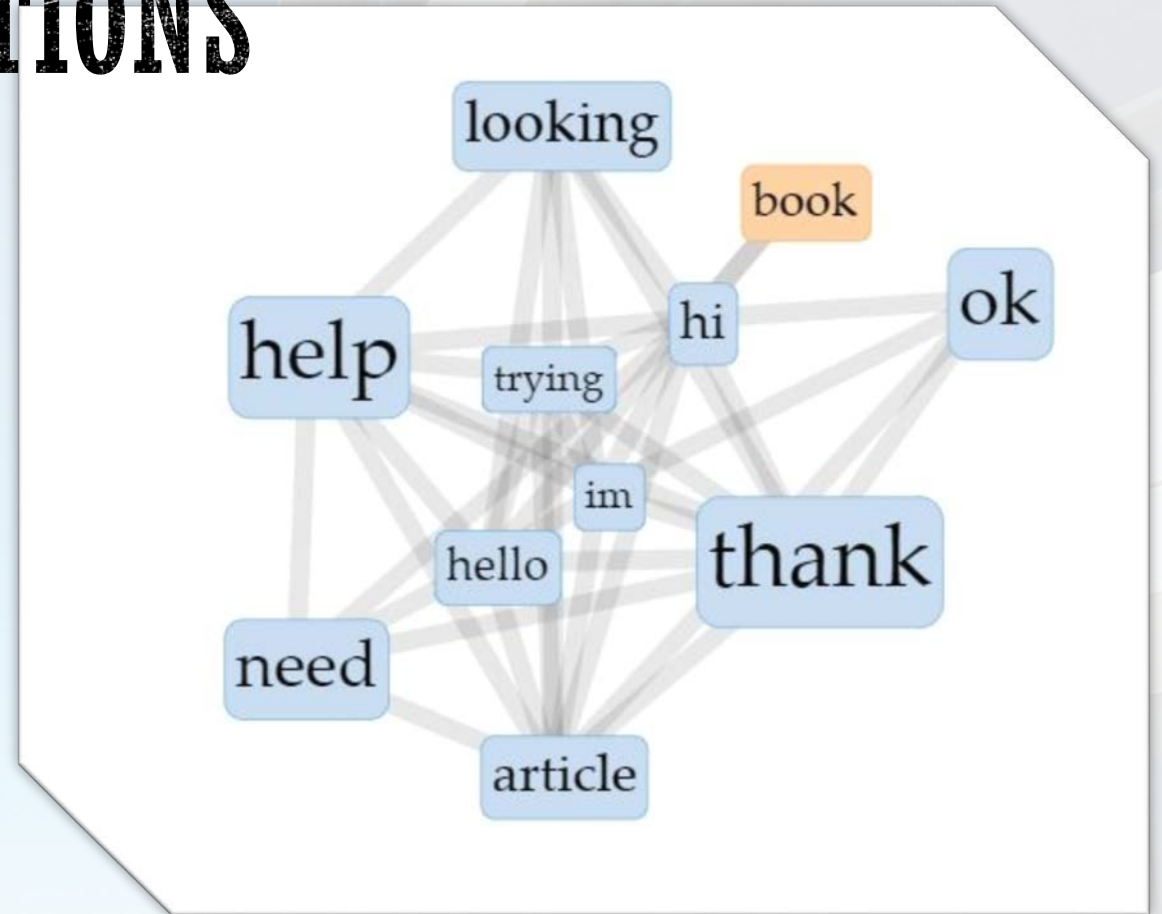
TOOLS AND METHODOLOGY

- **Lexos**
- **Voyant**
- **Subjectivity Lexicon**
- **LIWC2015**
- **Topic Modeling Tool**

1. We were supplied with anonymized copies of transcripts in the form of two distinct corpuses, "librarian" and "student"

2. We used Lexos and Voyant software to extract words and phrases from the chat transcripts and to establish word counts and frequencies for student vocabulary vs. librarian vocabulary

3. The Subjectivity Lexicon gave us a way to use Lexos again, by filtering for the designated sentiment word lists. Also used LIWC.

4. Input all transcripts (not separated- just one big file) into Topic Modeling Tool

TOP 100 WORDS (WEIGHTED)

Students

Librarians
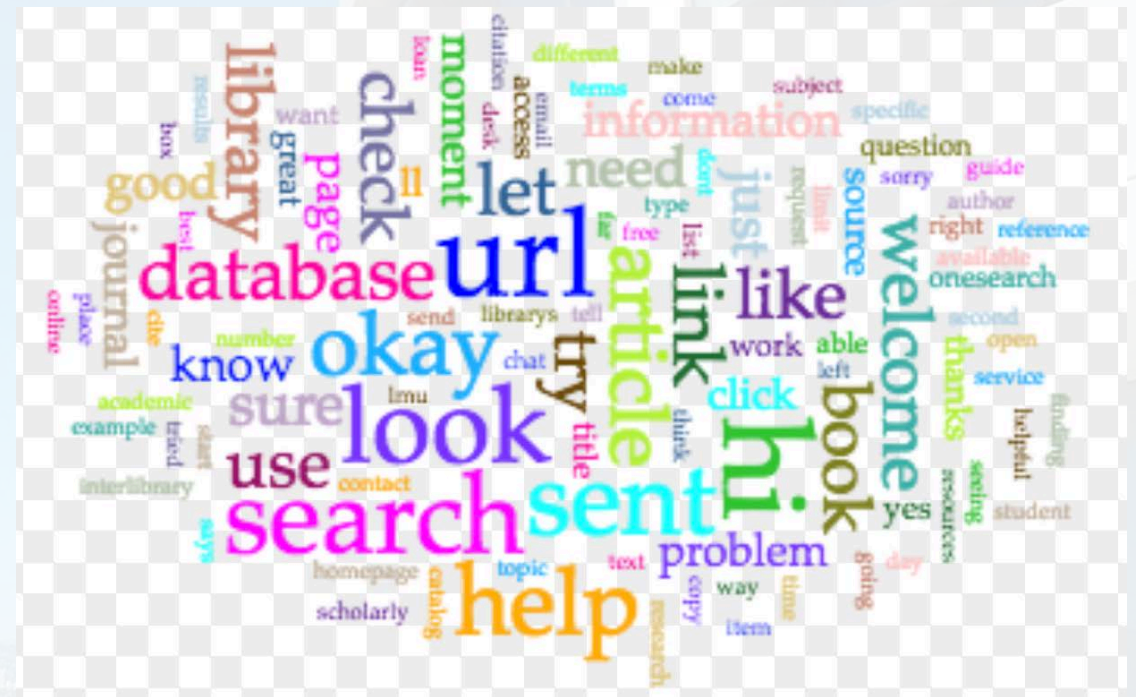
STUDENTS- TOP 100 WORDS (UNWEIGHTED)

LIBRARIANS- TOP 100 WORDS (UNWEIGHTED)

STUDENT RESEARCH VOCABULARY: TOP 23 TERMS

Research Term	Students- Times Mentioned	Librarians- Times Mentioned
1. Article	5039	5309
2. Book	3644	3519
3. Journal	1506	1931
4. Link	1437	3968
5. Database	1304	4233
6. Online	1198	1202
7. Topic	728	1023
8. Cite	667	301
9. Scholarly	539	640
10. Primary	530	542
11. Website	515	463

STUDENT RESEARCH VOCABULARY: TOP 23 TERMS

Research Term	Students- Times Mentioned	Librarians- Times Mentioned
12. Citation	409	698
13. eBook	346	270
14. PDF	322	427
15. Peer Reviewed	247	336
16. Newspaper / News	222	418
17. Reference –	220	560
18. Full Text	205	827
19. Chapter	141	241
20. Abstract	131	105
21. Popular	98	107
22. Background	86	104
23. Index	65	119

TOP RESOURCES

Tool/Resource	Students- Times Mentioned	Librarian- Times Mentioned
Ebsco/Ebscohost	317	188
Catalog/Linus	266	1168
MLA	253	258
Google	247	371
OneSearch	224	1519
APA	213	183
JSTOR	210	348
Illiad (interlibrary loan)	175	1085
Proquest	163	313
ERIC	156	226
Encyclopedia	124	284
LibGuide	113	466
Google Scholar	104	184
ATLA	85	186
Academic Search Complete	48	307
PsycINFO	45	198

TOP SUBJECTS

Subject	X Mentioned by Students	X Mentioned by Librarians
History	410	417
Education	322	364
Theology or Religion	305	376
Art	246	144
Business	204	530
Psychology	181	186
Gender	179	107
Science	141	239
Literature	139	212
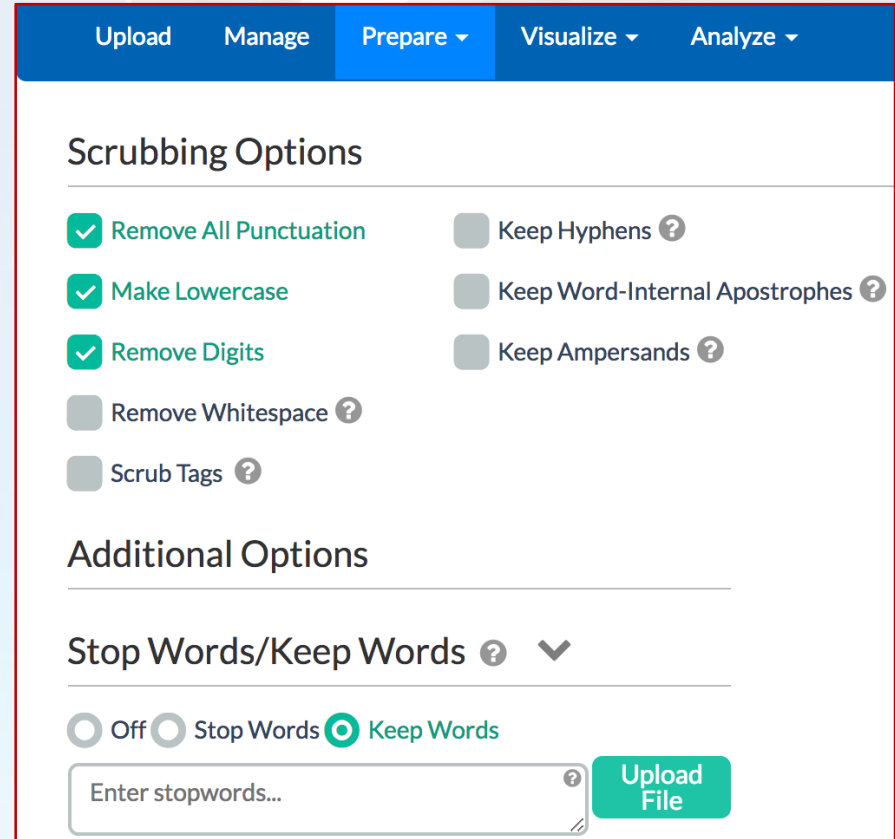
SENTIMENT ANALYSIS

- Bing Liu: "…the field of study that analyzes people's opinions, sentiments, appraisals, attitudes, and emotions toward entities and their attributes expressed in written text."
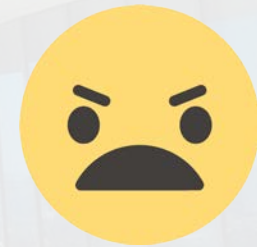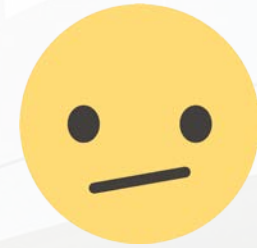
SENTIMENT ANALYSIS

- We applied sentiment analysis using *The Subjectivity Lexicon* and compared students to librarians.

- In Lexos: uploaded positive and negative terms as "keep words" in Lexos

- Identified percentages within student and librarian chats

[👍] SENTIMENT BY NUMBERS

Positive Words

- 3% of all librarian words in chat are positive

- 4% of all student words in chat are positive

Negative Words

- 1% of all librarian words in chat are negative

- 2% of all student words in chat are negative

NOTEWORTHY NEGATIVE WORDS-MORE THAN 2X DIFFERENCE

Negative Word	Student Word Count	Librarian Word Count
problem	355	2279
trying	1548	684
Limit/limited	65	712
down	196	594
trouble	477	142
little	134	338
hard	249	110
narrow	89	213
lost	139	55
rhetorical	113	45
afraid	7	111
unable	92	37
cross	10	61
tricky	5	61
busy	6	56

Positive Word	Student Word Count	Librarian Word Count
welcome	4	3455
sure	902	3271
good	1242	2488
might	283	2252
able	767	1535
right	681	1510
free	131	546
luck	100	543
perfect	531	77
top	50	355
awesome	353	55
glad	8	345
happy	59	316
above	31	218
useful	70	217
hope	93	215
fine	195	73
amazing	46	2

UNIQUE NEGATIVE WORDS PER CATEGORY ACROSS GROUPS

Students	Librarians
Nah (14)	Hesitate (23)
Nervous (8)	Warn/warned (6)
Nowhere (5)	
Suck (5)	
Bored (4)	
Crap/crappy (4)	
Rude (4)	
Hinder (4)	
Undesirable (4)	

UNIQUE POSITIVE WORDS PER CATEGORY ACROSS GROUPS

Students	Librarians
Beautiful (12)	None
Gracias (10)	
Impressed/impressive (8)	
Grateful/gratitude (7)	
Alrighty (6)	

UNIQUE WORDS PER CATEGORY ACROSS GROUPS

Verbs

Students	Librarians
Broke (63)	Exclude (6)
	Replacing (4)

Affiliation

Students	Librarians
Roommate (7)	Liaisons (14)
Girlfriend (4)	Consortium (14)

UNIQUE INFORMAL LANGUAGE FROM STUDENTS

Swear Words

Dang (6)
Damn (4)
Freaking (3)
Fuck (2)
Shit (2)

Netspeak or Informal

Sooo (up to 14 o's) (26)
Plz (9)
Nvm (6)
Dude (4)
Thanx (4)
LMAO (4)

TOPIC MODELING

- Topic Modeling is a method for finding and tracing clusters of words (called "topics") in large bodies of text.

- A topic model is a simplified representation of a collection of documents (in our case, 6 semesters of chat transcript files).

- Topic modeling software identifies words with topic labels, such that words that often show up in the same document are more likely to receive the same label.

- It can identify common subjects in a collection of documents that have similar meanings and associations.

TOPIC MODELING TOOL

- **MALLET**, a package of Java code, uses the command line.

- **Topic Modeling Tool** adapted from the original Mallet tool.

- TMT implements MALLET in a graphical user interface (GUI), meaning you can plug files in and receive output without entering a line of code. Yay!

- Helps minimize coding bias.

- Program does the grunt work.

TOPIC MODEL RESULTS

- General Reference Questions

- Navigation / Finding Questions

- Access Barriers & Policy Questions

- Technical and Specific Resource Questions

- Specific Subjects and Topics

- Meta / Chat Related Questions and Referrals

GENERAL REFERENCE QUESTIONS

Citation in different formats/citation styles (321)	Tangentially related- research questions (54)
Using the Journal List looking for a specific citation (265)	Tangentially related- keyword and search construction tips esp. with the words relationship or benefit or support (52)
Finding full text of an article (239)	Background information/tertiary sources (47)
Is there an ebook (189)	Finding statistics (32)
Primary source questions- is this primary, how can I find (146)	Tangentially related- finding articles (31)
Academic journals and article searching (125)	Citing authors or using footnotes, creating Annotated Bibliography (30)
Tangentially related- help starting research for books and/or articles (92)	Citation questions, particularly about dates of publication (28)
Finding sources on a topic (91)	Find this article/do we have this journal (25)
Tangentially related- help with searching databases (64)	Tangentially related- questions about finding books and articles (22)
Popular/scholarly and encyclopedias (60)	Reference interview searching and changing keywords (20)

NAVIGATION/FINDING QUESTIONS

Where could I find this call number (198)	Variations on "I'm trying to find" and "I'm having trouble finding" (88)
Interlibrary loan and ILLiad accounts (185)	Reserves, e-reserves, other Blackboard questions & demo class chats (80)
"library website" and "search box" questions (often related to Onesearch, library catalog) (173)	Access to newspapers (New York Times, LA Times, Wall Street Journal) (76)
Read Online and Request article links (131)	LMU student and faculty services questions, various forms and privileges (74)
Using filters and limiters (109)	Questions about links- Blackboard, permalink, ebook, DOI (68)
How do I find this database (105)	Requests for various media items (66)
Is this available in print and is this available online (98)	Research consultation requests (25)
Scholarly Peer Review Checkbox (97)	

ACCESS BARRIERS & POLICY QUESTIONS

Access Barriers

Issues accessing databases using MyLMU username and password/remote access (311)
Looking for textbooks (186)
Placing a hold or requesting from the basement (174)
Access to databases- at all or off campus (98)
Saving an ebook to your computer, mainly Ebscohost (71)

Policy

How many books can undergraduate check out (265)
Library hours (215)
Questions about fines and renewals (121)
Can I get a library card (various nonstudents) (102)
Requesting books and arrival time (71)

TECHNICAL AND SPECIFIC RESOURCE QUESTIONS

Technical

Technical equipment and software questions (111)
Reserving a group study room and other issues with study rooms (99)
Issues with the chat box (78)
Printing costs and how to add funds (77)

Specific Resources

OneSearch+ (186)
Suggesting LibGuides (117)
Using Google Scholar- articles and cited by feature (25)

SPECIFIC SUBJECTS/TOPICS

Keyword help on topic related to minimum wage (106)	Law research and LexisNexis (33)
Business research questions, usually industry (97)	Controversial issues research (32)
Theology research questions (91)	More Education topics (32)
Sociology questions mainly around racial discrimination and diversity (91)	Communication topics (30)
Education research questions (81)	Tangential: Sin, silence, suffering, & confession in Psalm 32 (26)
Psychology research questions (78)	Mixed topics re Greek, music, and composers (25)
Article requests with a focus on public health, masculinity, and race/ethnicity (39)	Physical therapy and injuries (24)
Research questions involving women (39)	Economic research and churches or traditions (24)
Minority group representation and discrimination (37)	Partially related- world war two topics (23)
Scientific topics, especially rhizobacteria (37)	Health sciences related topics (23)
Medical articles, art therapy, and other science (37)	Tangential: research about college students (23)
Literature criticism & MLA Bibliography (36)	Tangential: book searches about places (23)
Race and class in mostly Mexico (36)	Tangential: places and sourcebooks (19)
Film topics (33)	Tangential: people or related to subject headings (18)

META/CHAT RELATED QUESTIONS AND REFERRALS

Meta

Greetings and signoffs (191)
Tangentially related- nonLMU staff helping w/ local inquiries (121)
Unrelated- "no problem" (115)
Tangential relationship around the word information (107)
Tangentially related under "I don't know" (68)
Hey and random chatting/spam chat (61)
Shift change new person staffing chat (27)
Tangentially related- chats with foreign languages in them (7)

Referrals

Local question referral from AJCU librarians (187)
Forwarding to LMU library staff for follow-up (79)
Questions about Special Collections (54)
Referrals to call departments on campus or go to Circulation desk (53)
Academic Resource Center referrals (37)

CONCLUSIONS

- Librarians contribute 59% more to each chat conversation than students and are wordier

- Overall, students display slightly more affective processes than librarians (4.79 words per chat vs. 3.09 words per chat)
  - Students display slightly more positive emotions
  - Students also display slightly more negative emotions

- Overall, students and librarians use a similar amount of informal language (2.18 words per chat vs. 2.04)

- Chat maintains a polite atmosphere- very little swear words used by anyone

CONCLUSIONS

- Students use vendor name "Ebsco/Ebscohost" rather than individual database name (e.g. Academic Search Complete)

- Sentiment dictionaries are not designed for a library context, so have limited usefulness

- Tools used were free, but required follow-up with experts, and lots of time

- Algorithm in Topic Modelling Tool occasionally found peripheral relationships (e.g. chats w/ foreign languages, book searches about places, use of phrase "no problem")

Applications from What Was Learned

- Improving description of library resources in LibGuides with terminology students will recognize

- Improving positive atmosphere in chat by aligning provider vocabulary with terms favored by students (e.g. perfect, awesome, amazing)

- Frequent subjects included Business, Theology, Education, History, and Psychology (outreach implications to more departments)

- We will use frequent topics to improve library FAQs, navigation cues, and access issues

APPLICATIONS FROM WHAT WAS LEARNED

The library needs to look more carefully at problem areas that cut across library departments:

- Downloading or saving EBSCOhost eBooks
- Remote access to databases
- Better navigational help and directions in catalog, article linking, and building
- Collect more online textbooks
- Group study room issues
- Better tips for student use of permanent URLs
- Still too much library jargon (e.g. ILLiad, LINUS, "borrowing privileges")