

**Loyola Marymount University**

---

**From the Selected Works of Susan Gardner Archambault**

---

Spring April, 2018

# Algorithmic Accountability, AI, Transparency, & Text Analysis Assessment Panel

Susan [Gardner] Archambault

Alexander Justice, *Loyola Marymount University*



Available at: [https://works.bepress.com/susan\\_gardner/28/](https://works.bepress.com/susan_gardner/28/)

# TEXT ANALYSIS ASSESSMENT

Using quantitative text analysis to assess and  
improve a library service

[Susan.Archambault@lmu.edu](mailto:Susan.Archambault@lmu.edu)

[Alexander.Justice@lmu.edu](mailto:Alexander.Justice@lmu.edu)

Loyola Marymount University, Los Angeles



# CHAT REFERENCE

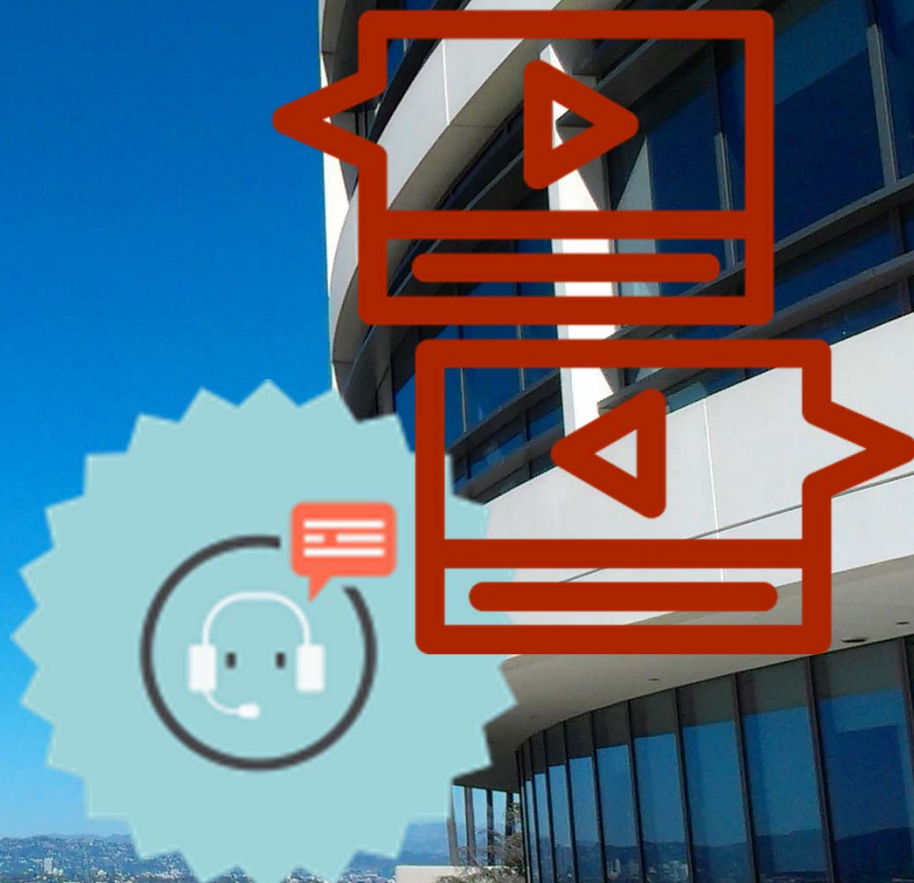
- Part of LMU library reference service since 2004
- We provide chat reference 24/7 by working with a consortium of most other Jesuit universities in US
- Our platform is libraryh3lp, a product of Nub Games, Inc.
- Libraryh3lp is set up to save transcripts of all chat sessions

The screenshot displays the LMU LA Academic Search Complete interface. The top navigation bar includes links for General Info, Services, Technology, Research, Collections, Happenings, FAQ, and My Account. Below this, a secondary bar offers options like New Search, Publications, Subject Terms, Cited References, and More. The main search area features a search bar with the query 'jodo shinshu' and buttons for Search and Clear. To the right of the search bar is a 'Library Logo' link. Below the search bar, there are sections for 'Refine Results' and 'Search Results: 1 - 10 of 23'. The 'Refine Results' section includes a 'Current Search' box with the query 'jodo shinshu', a 'Limit To' section with checkboxes for 'Full Text', 'References Available', and 'Scholarly (Peer Reviewed) Journals', and a 'Publication Date' range from 1990 to 2016. The 'Search Results' section lists two results. The first result is 'The Ritual Use of Music in US Jōdo Shinshū Buddhist Communities' by Mitchell, Scott A., published in Contemporary Buddhism, Nov 2014, Vol. 15 Issue 2, p356-372. The second result is '"All Beings Are Equally Embraced By Amida Buddha": Jodo Shinshu Buddhism and Same-Sex Marriage in the United States' by Wilson, Jeff, published in the Journal of Global Buddhism, 2012, Vol. 13, preceding p31-59. Both results include links for 'PDF Full Text' and 'Catalog Record'. On the right side of the page, there is a 'Ask an LMU Librarian' chat window with a text input field and a 'Type here to chat. Press ENTER to send.' prompt.



# THE DATA

- We used six semesters' worth of chat transcripts as a corpus for a variety of text analysis research.
- January 2014 to December 2016
- Over 10,000 conversations





# 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





# PREPARING THE CORPUS

- Removed metadata, punctuation, digits, stop words
- Transformed tokens (consolidate, lemmas) e.g. e-book, ebook, ebooks: ebook

**Lexos {loader}** An Integrated Lexomics Workflow

Workspace

Reset



Upload

Manage

Prepare ▾

Visualize ▾

Analyze ▾



## File Upload

Max. file size: 250 MB -- Formats Supported: .txt, .html, .xml, .sgml, and .lexos ?

Browse

Ready For Files To Upload

Or drop files here



# LEXOS

## Lexos {scrubber} An Integrated Lexomics Workflow

[Workspace](#)[Reset](#)[Upload](#)[Manage](#)[Prepare](#)[Visualize](#)[Analyze](#)

### Scrubbing Options

- |  |   |
|--|---|
| <input checked="" type="checkbox"/> Remove All Punctuation | <input type="checkbox"/> Keep Hyphens ?                   |
| <input checked="" type="checkbox"/> Make Lowercase         | <input type="checkbox"/> Keep Word-Internal Apostrophes ? |
| <input checked="" type="checkbox"/> Remove Digits          | <input type="checkbox"/> Keep Ampersands ?                |
| <input type="checkbox"/> Remove Whitespace ?               |   |
| <input type="checkbox"/> Scrub Tags ?                      |   |

### Additional Options

#### Stop Words/Keep Words ? ▾

☒ Off ☐ Stop Words ☐ Keep Words

[Upload File](#)

#### Lemmas ? ▾

[Upload File](#)

#### Consolidations ? ▾

[Upload File](#)

#### Special Characters ? ▾

- Select Built-In Option - ▾

[View Pre-Defined Rule Sets](#)[Upload File](#)[Preview Scrubbing](#)[Apply Scrubbing](#)[Download Scrubbed Files](#)

### Previews of Documents

LEXOS-

NOSTOPWORDSLEMMAScombined\_chatfiles

hi wondering guys rent financial calculator hi moment look financial aid calculator no financial calculator finance class texas ba not seeing calculators list checkout suggest check circulation desk let number thanks thanks hi look article anchor bi...able history libguide lmu librarys webpage linked sent url cambridge oxford online might good information not want check academic database list topic specific cambridge oxfords online reference materials just clarify look materials about s correct



## Original Chat File

Removed LMU student ID numbers and student email addresses,  
librarian identifiers

```
08:31PM 03053321684380181924925096: Hello. I am trying to search for art therapy articles about defense mechanisms in incarcerated teens but am not finding anything. Just want to check if I'm doing it the right way. On the psych info data base, I'm typing adolescent art therapy in the first box, and defense mechanisms in the second. Any ideas?
08:33PM 03053321684380181924925096: Ok I'll try that. Other than that I'm using it correctly though?
08:34PM 03053321684380181924925096: Ok I hope so. I'll try again.
08:38PM 73592669818591801761253655: worldcat has pointed me to an online resource; how do I access it? I searched AUTHOR Segrera, Alberto S, LANGUAGE Spanish, and the first of the four items is the one I am interested in.
08:54PM 73592669818591801761253655: Enfoque centrado en la persona : bibliografía en español 1997nOCLC: 757971878
08:55PM 73592669818591801761253655: my e-mail address is:
08:55PM 73592669818591801761253655: EMAIL-ADDRESS-REMOVED
```

## Scrubbed Chat File

Stripped all numbers, timestamps, replaced URLs with "sent URL." Removed stop words  
determined by us, then consolidated lemmas.

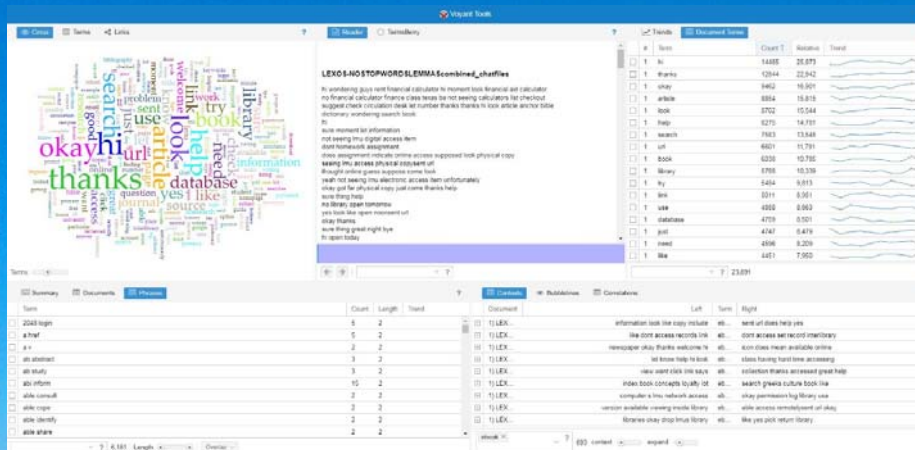
```
hi try search art therapy article about defense mechanisms incarcerated teens not finding just want check doing right way psych information data base typing adolescent art therapy box defense mechanisms second ideas okay ll try use correctly okay hope ll try worldcat pointed online resources access search author segrera alberto s language spanish item interested enfoque centrado en la persona bibliografía en español noclc e-mail address
```



# EXPLORING WORD FREQUENCIES

Voyant

Lexos



Tokenize

2

-gram

☒ by Tokens

☐ by Characters

Normalize

☐ Proportional Counts

☒ Raw Counts

☐ Weighted Counts (TF-IDF)

Assign Temporary Labels

Culling Options

☒ Most Frequent Words

Use top 

100

 terms

☐ Culling

Download CSV

Download TSV

Search:









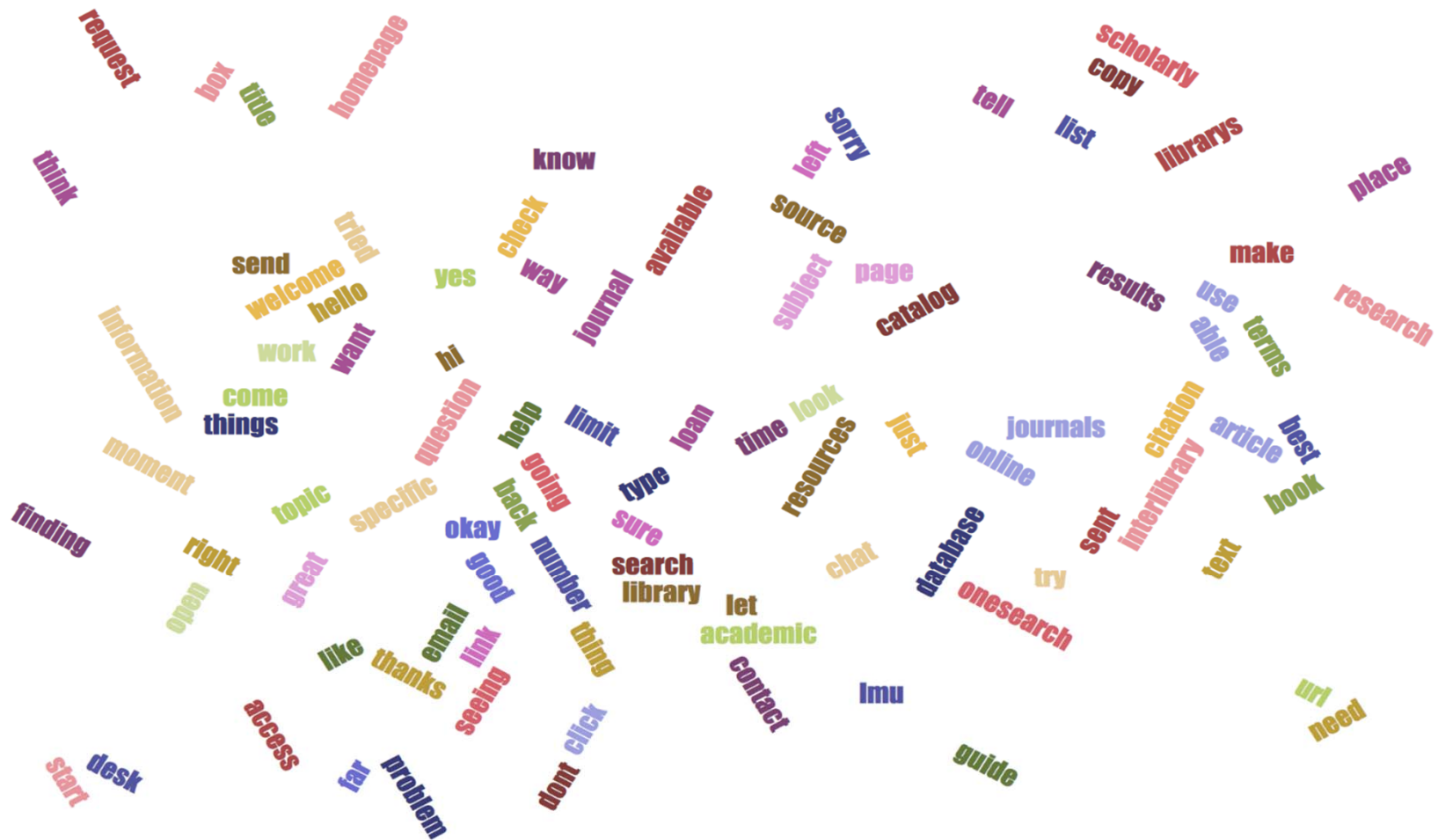


# STUDENTS- TOP 100 WORDS (UNWEIGHTED)





# 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."





**Main**  
MPQA Home

**Corpora**  
News, debates, etc.

**Lexicons**  
Subj. clues, etc.

**Annotation**  
GATE, MPQA, gfbf

**OpinionFinder**  
Subjectivity detector

## Lexicons

Subjectivity Lexicon  
[Subj. Sense Annotations](#)  
[Arguing Lexicon](#)  
[+/-Effect Lexicon](#)

## Subjectivity Lexicon

Made available under the terms of [GNU General Public License](#). They are distributed without any warranty.

**The Subjectivity Lexicon** (list of subjectivity clues) that is part of OpinionFinder is also available for separate download. These clues were compiled from several sources (see the enclosed README). This is the version of the lexicon used in:

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann (2005). [Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis](#). *Proc. of HLT-EMNLP-2005*.

To access the **MPQA Subjectivity Lexicon**, please fill out the following form. We respect your privacy and will not use your information for any purpose other than to inform you of updates and to assess interest in the resource.

Name:

Institution:

E-Mail:

contact: [mpqa.project@gmail.com](mailto:mpqa.project@gmail.com)

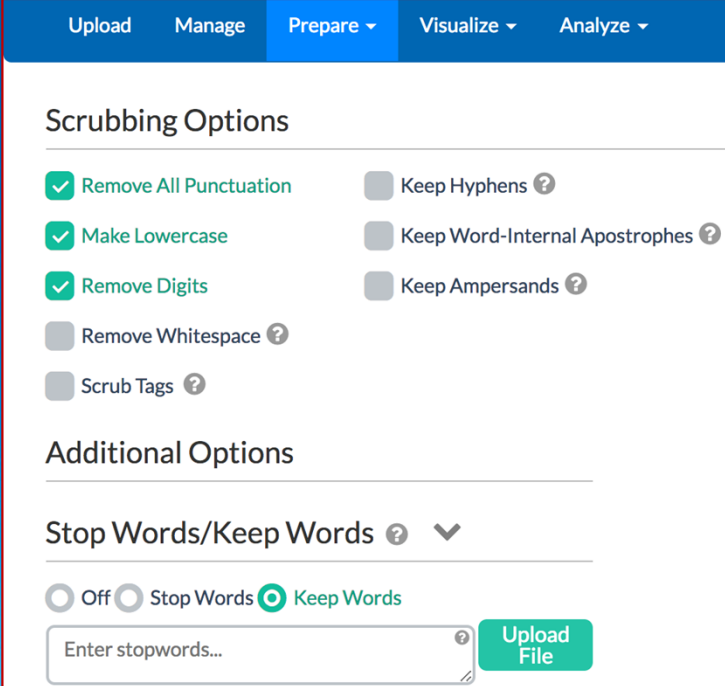
[\[nlp\]](#) [\[cs\]](#) [\[pitt\]](#)





# 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



The screenshot shows the Lexos web application interface. The top navigation bar includes 'Upload', 'Manage', 'Prepare' (selected), 'Visualize', and 'Analyze'. The 'Prepare' tab is active, displaying 'Scrubbing Options' and 'Additional Options'.

**Scrubbing Options**

<input checked="" type="checkbox"/> Remove All Punctuation	<input type="checkbox"/> Keep Hyphens ?
<input checked="" type="checkbox"/> Make Lowercase	<input type="checkbox"/> Keep Word-Internal Apostrophes ?
<input checked="" type="checkbox"/> Remove Digits	<input type="checkbox"/> Keep Ampersands ?
<input type="checkbox"/> Remove Whitespace ?	
<input type="checkbox"/> Scrub Tags ?	

**Additional Options**

Stop Words/Keep Words ? ▾

☐ Off ☐ Stop Words ☒ Keep Words

Enter stopwords... ?





# [ ] 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



# LINGUISTIC INQUIRY AND WORD COUNT (LIWC2015)

## Welcome to LIWC



### Analyze Text

See how text measures up along LIWC's dimensions.



### Categorize Words

Break down each word and find the categories it applies to.



### Color-Code Text

Highlight words found in LIWC's dictionary.



### Category Options

Choose which dimensions to use during analysis.



### Help

Learn more about what LIWC does and how to use it.

#### YOUR SETTINGS

##### *Dictionary*

Internal Dictionary 2015

##### *Segmentation*

No segmentation

##### *All categories on*

NO



# COMPARISON ACROSS GROUPS:

average word per category in each chat

Filename	Segment	WC	WPS	Dic	negate	verb	affect	posemo	negemo	anx	anger	sad
all-students.txt	1	793988	18.51	63.72	1.22	13.79	4.79	4.15	0.61	0.09	0.10	0.17
all-librarians.txt	1	1263621	20.79	67.10	0.82	12.00	3.09	2.53	0.55	0.04	0.06	0.14

Filename	Segment	affiliation	reward	risk	informal	swear	netspeak	assent	nonflu	filler
all-students.txt	1	2.05	1.22	0.26	2.18	0.01	0.58	1.36	0.25	0.02
all-librarians.txt	1	2.43	1.09	0.33	2.04	0.00	1.32	0.62	0.20	0.00



# 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)	
Almighty (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

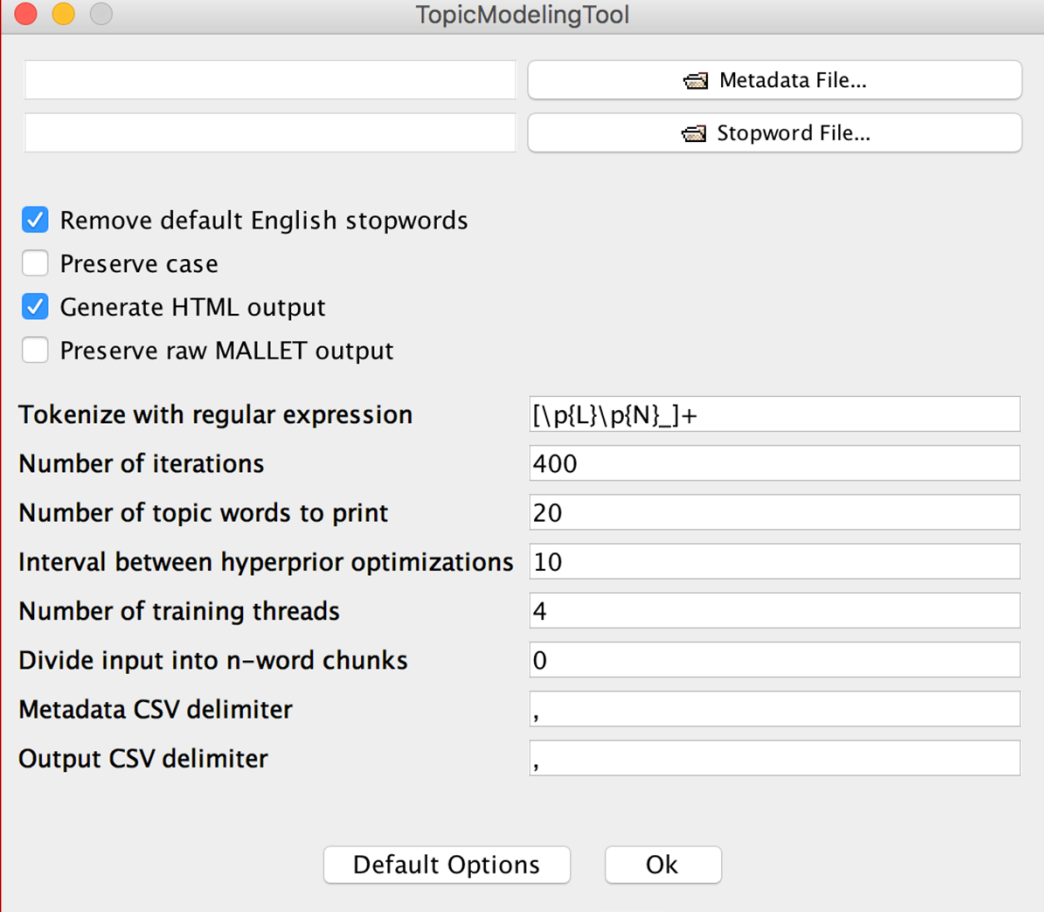
- MALLET, a package of Java code, uses the command line.
- Topic Modeling Tool adapted from the original Mallet tool.
- TMT implements MALLET in a graphical user interface (GUI), meaning you can plug files in and receive output without entering a line of code. Yay!
- Helps minimize coding bias.
- Program does the grunt work.





# HOW IT WORKS

- csv or html output
- You control number of topics it delivers & number of words in each topic cluster
- The topic model learns from patterns of words that tend to co-occur within documents

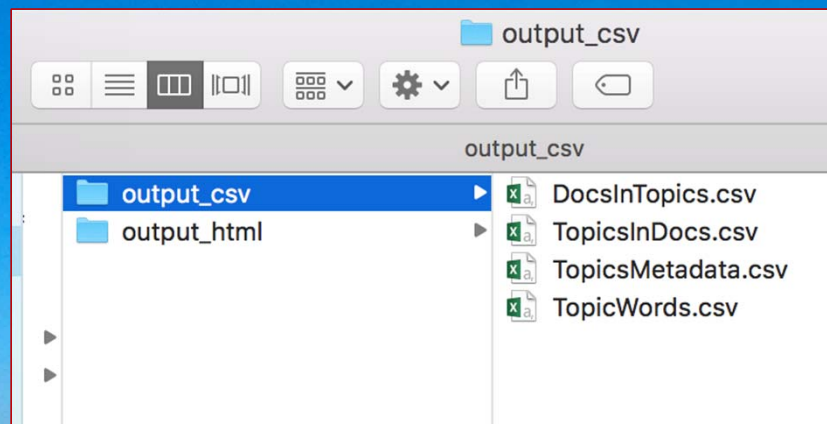


The screenshot shows a window titled "TopicModelingTool" with a light gray background. At the top, there are two empty text input fields. To the right of these fields are two buttons: "Metadata File..." and "Stopword File...". Below these are four checkboxes: "Remove default English stopwords" (checked), "Preserve case" (unchecked), "Generate HTML output" (checked), and "Preserve raw MALLET output" (unchecked). Further down, there are several labeled text input fields: "Tokenize with regular expression" with the value "[\\p{L}\\p{N}\_]+", "Number of iterations" with the value "400", "Number of topic words to print" with the value "20", "Interval between hyperprior optimizations" with the value "10", "Number of training threads" with the value "4", "Divide input into n-word chunks" with the value "0", "Metadata CSV delimiter" with the value ",", and "Output CSV delimiter" with the value ",". At the bottom of the window, there are two buttons: "Default Options" and "Ok".

Tokenize with regular expression	[\\p{L}\\p{N}_]+
Number of iterations	400
Number of topic words to print	20
Interval between hyperprior optimizations	10
Number of training threads	4
Divide input into n-word chunks	0
Metadata CSV delimiter	,
Output CSV delimiter	,

# TOPIC MODELING TOOL OUTPUT

Topic Id	Top Words...
	cite citation style mla apa citing format chicago page citations guide source examples owl website author put purdue manual cited include quote question online correct reference text resource guidelines in-text site part works bibliography numbers end section formatting double scroll 2 paragraph instructor work give retrieved helpful follow confused write sentence



TOPIC : theology religion atla find bible catholic dance studies church christian religious books theological paper christianity commentaries source sources yoga testament ...

top-ranked docs in this topic (#words in doc assigned to this topic)

1. (132) [2016-04-09.224803.xml.txt](#)
2. (94) [2015-03-17.054517.xml.txt](#)
3. (93) [2015-11-25.023028.xml.txt](#)
4. (86) [2015-04-06.021401.xml.txt](#)
5. (77) [2014-02-23.061242.xml.txt](#)
6. (61) [2014-02-23.051805.xml.txt](#)
7. (60) [2015-04-04.192425.xml.txt](#)
8. (59) [2014-10-28.013812.xml.txt](#)
9. (57) [2015-11-26.015742.xml.txt](#)
10. (56) [2014-10-29.162333.xml.txt](#)
11. (55) [2014-03-04.230950.xml.txt](#)
12. (53) [2016-03-13.165751.xml.txt](#)
13. (52) [2016-11-21.220637.xml.txt](#)
14. (45) [2016-10-13.020113.xml.txt](#)
15. (44) [2016-04-18.195801.xml.txt](#)
16. (44) [2015-03-25.140741.xml.txt](#)
17. (42) [2016-11-10.205630.xml.txt](#)
18. (42) [2014-11-05.221003.xml.txt](#)
19. (42) [2014-10-04.175129.xml.txt](#)
20. (41) [2014-10-22.224901.xml.txt](#)



# 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”)





# THANK YOU

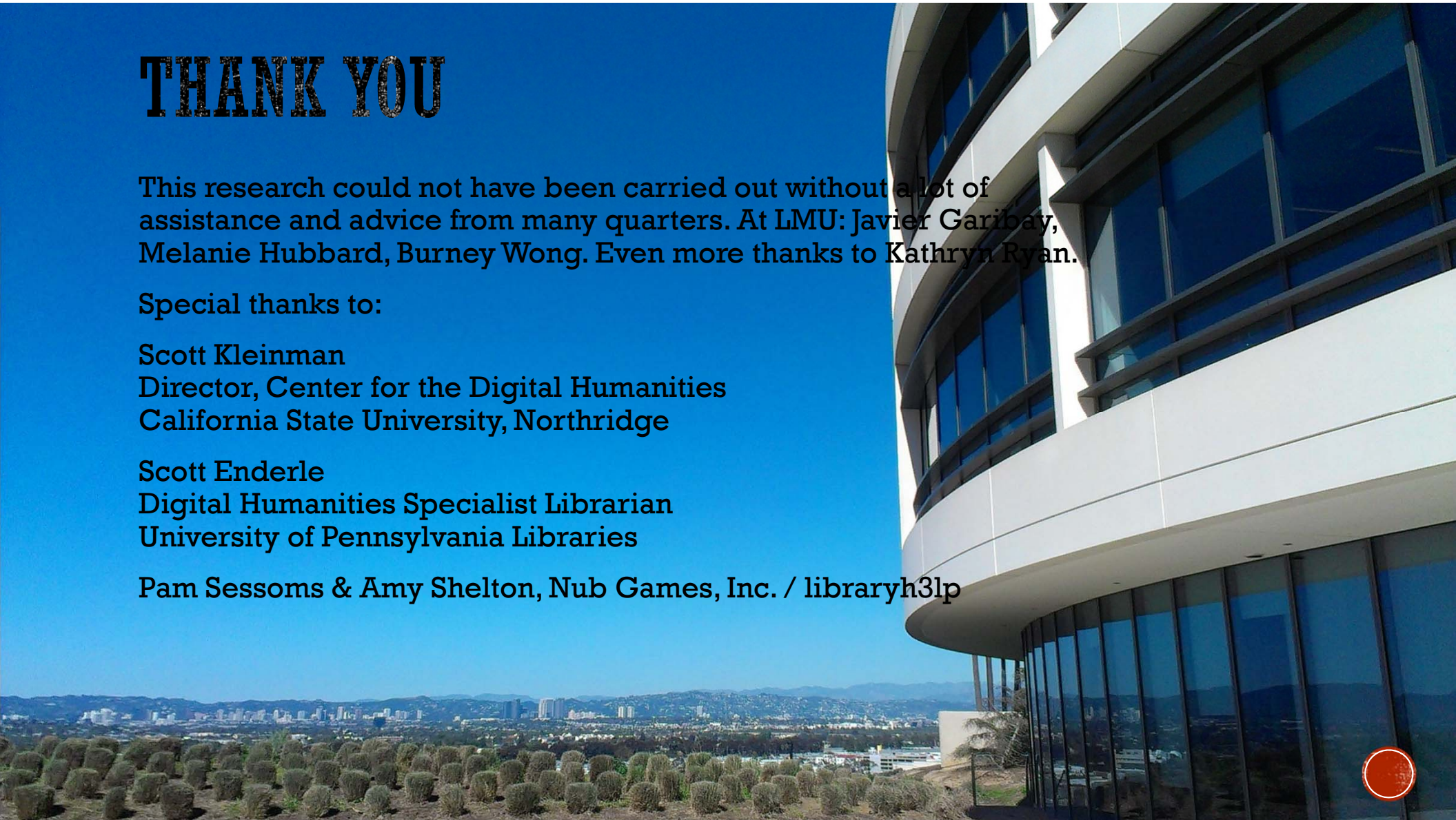
This research could not have been carried out without a lot of assistance and advice from many quarters. At LMU: Javier Garibay, Melanie Hubbard, Burney Wong. Even more thanks to Kathryn Ryan.

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](http://libraryh3lp.com)





# VOCABULARY

- Algorithm- a process a computer follows to solve a problem, creating an output from a provided input
- Lexos- a web-based tool used for pre-processing, analysis, and visualization of digitized texts. See <http://lexos.wheatoncollege.edu>
- Linguistic Inquiry and Word Count (LIWC)- a propriety database consisting of a lot of categorized regular expressions; counts the percentage of words that reflect different emotions. See <https://liwc.wpengine.com/>
- Machine Learning- a process that gives computers the ability to learn without being explicitly programmed based on algorithms that can learn from and make predictions on data
- Mental Model- explanation of someone's thought process
- Natural Language Processing (NLP)- using computers to understand human-language text
- Sentiment Analysis- using computers to systematically identify attitudes or emotions present in text
- Stop words- frequently used words (such as "the," "and," "if") that are often removed from text before performing analysis
- Subjectivity Lexicon- The MPQA (Multi-Perspective Question Answering) Subjectivity Lexicon is maintained by Theresa Wilson, Janyce Wiebe, and Paul Hoffmann (Wiebe, Wilson, and Cardie 2005). See [http://mpqa.cs.pitt.edu/lexicons.html\\_lexicon/](http://mpqa.cs.pitt.edu/lexicons.html_lexicon/)
- Text Analysis- a form of data mining, using computer-aided methods to study textual data
- Text Corpus- a "corpus" of text is the individual research text dataset
- Tokenization- breaking text into pieces called tokens. Often certain characters (punctuation marks, numbers) are discarded in the process.
- Topic Modeling- a method of using statistical models for discovering the abstract "topics" that occur in a collection of documents
- Topic Model Tool- A point-and-click tool for creating and analyzing topic models produced by MALLET. See <https://github.com/senderle/topic-modeling-tool>
- Voyant- a tool that can create many types of visualizations such as word clouds, bubble charts, and word trees. See <https://voyant-tools.org/>
- Word Cloud/tag- a graphical representation of word frequency

