Comprehending the Semantic Essence of Text: Leveraging SEANCE for Textual Analysis

Instructor: Lingzi Hong & Xiaoying Song

Presenter Profile

- Presenter: Lingzi Hong
- Assistant Professor in Data Science at the University of North Texas
- Research Interest:
  - Computational Linguistics
  - Human-Centered Computing
  - Social Media Analysis
  - Data Literacy

Presenter Profile

- Presenter: Xiaoying Song
- PhD Student in Information Science at the University of North Texas
- Research Interest:
  - Large Language Models
  - Online Misbehavior
  - Counter Speech
  - User Information Behavior

Outline

- **Theoretical Part (35 minutes)**
  - Natural Language Processing
  - Text Analysis
  - SEANCE for Text Analysis
  - Results Interpretation
- **Demonstrations (15 minutes)**
  - Setting Up SEANCE
  - Running SEANCE
  - Statistical Test
- **Q&A (10 minutes)**

01

Theoretical Part

Natural Language Processing

Definition

Natural language processing (NLP) is the ability of a computer program to understand human language as it's spoken and written -- referred to as natural language.

Application

Several NLP tasks break down human text and voice data in ways that help the computer make sense of what it's ingesting, such as sentiment analysis, named entity recognition, natural language generation.

Applications of NLP in Libraries

Text Mining and Analysis

- **Scenario:** The City Library receives a lot of feedbacks about the service through emails, social media, and surveys. Librarians hope to understand user needs and gauge satisfaction with existing materials and services.
- **Methods:** Automatic analysis of sentiment and entities through NLP tools.
- **Example insights:** Users desire more e-book collection; Users are not satisfied with certain activity for children; Users would like to see more new books; etc.

Applications of NLP in Libraries

Digital Archiving and Preservation

- **Scenario:** The Heritage Library has a significant collection of historical texts, including handwriting manuscripts, old newspapers, and government documents. The library needs to preserve these invaluable resources and make them accessible.
- **Methods:** Extraction of text from images. Automatic analysis of entities through NLP tools.
- **Example Results:** NLP tools automatically transcribe the text from scanned images. These transcribed texts are then annotated using NLP techniques to identify and tag key entity information, such as dates, names, and places.

Text Analysis

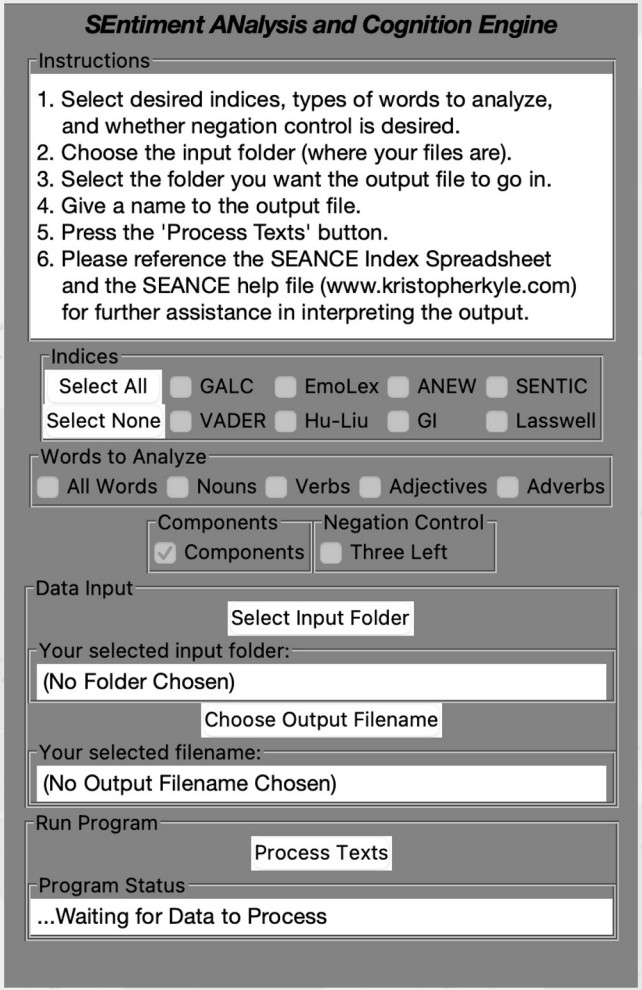
- Text analysis, a branch of Natural Language Processing (NLP), involves the automated processing of unstructured text to extract and categorize pertinent information using various techniques.

- These techniques include topic extraction, sentiment analysis, aspect classification, and named entity extraction, among others.

Source:https://medium.com/@aishwaryaprasad.ca/text-analysis-9eee8ade9754

Examples

- **App Review: "I absolutely loved the smooth interface and easy navigation of this app!"**
  - Sentiment: Positive
  - Emotion: Joy
  - Named Entity: App
  - Topic Category: Technology/Product Review
- **Social Media Feedback: "The Wi-Fi at City Library is so slow. It's frustrating trying to get any work done. #librarywoes!"**
  - Sentiment: Negative
  - Emotion: Frustration
  - Named Entity: City Library
  - Topic Category: Library Complaint



SEANCE

- SEANCE is a sentiment analysis tool that relies on a number of preexisting sentiment, social-positioning, and cognition dictionaries.
- SEANCE contains a number of predeveloped word vectors developed to measure sentiment, cognition, and social order. These vectors are taken from freely available source databases
- SEANCE is written in Python but is implemented in a way that requires little to no knowledge of programming, and it can be started by simply double-clicking the SEANCE icon.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2017). Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior research methods*, *49*, 803-821.

Source Database

GI	Seventeen semantic categories: semantic dimensions, pleasure, overstatements, institutions, roles, social categories, references to places, references to objects, communication, motivation, cognition, pronouns, assent and negation, and verb and adjective types.
Lasswell	Nine semantic categories: power, rectitude, respect, affection, wealth, wellbeing, enlightenment, and skill.
GALC	Emotion: 36 specific emotions and two general emotional states (positive and negative).
ANEW	Effective norms for valence, pleasure, arousal, and dominance.
EmoLex	Lists of words and bigrams that evoke particular emotions (e.g., anger, anticipation, disgust, fear, joy, sadness, surprise, and trust).
SENTIC	Emotional dimensions (sensitivity, aptitude, attention, and pleasantness)
VADER	A rule-based sentiment analysis system developed specifically for shorter texts found in social media contexts.
Hu–Liu	Polarity list developed specifically for product reviews and social texts.

Indices Examples

- **Arousal, Arousal_nwords:** These indices measure the level of arousal, which refers to the intensity of emotion provoked by words. High arousal words typically trigger a strong emotional response (e.g., excitement, alarm).
- **Valence, Valence_nwords:** Valence indices measure the positive or negative sentiment conveyed by words. Positive valence indicates pleasantness, while negative valence points to unpleasantness.
- **Longing_GALC, Lust_GALC, Anger_GALC**
  - Longing: Reflects a yearning or desire for something.
  - Lust: Associated with intense desire, often in a sexual context.
  - Anger: Measures expressions of irritability, frustration, and rage.

Indices References

	Index	Variable description	Variable Category	examples (not POS specific)
1	Arousal	Arousal	Arousal	Norms
2	Arousal_nwords	Arousal	Arousal	Norms
3	Dominance	Dominance	Dominance, respect, money, and power	Norms
4	Dominance_nwords	Dominance	Dominance, respect, money, and power	Norms
5	Valence	Valence	Valence/polarity	Norms
6	Valence_nwords	Valence	Valence/polarity	Norms
7	Longing_GALC	Longing	Arousal	crav*, daydream*, desir*, fanta*, hanker*
8	Lust_GALC	Lust	Arousal	carnal, lust*, climax, ecsta*, orgas*
9	Anger_GALC	Anger	Negative Emotion words	anger, angr*, cross*, enrag*, furious
10	Anxiety_GALC	Anxiety	Negative Emotion words	anguish*, anxi*, apprehens*, diffiden*, jitter*
11	Boredom_GALC	Boredom	Negative Emotion words	bor*, ennui, indifferen*, languor*, tedi*
12	Contempt_GALC	Contempt	Negative Emotion words	contempt*, denigr*, deprec*, deris*, despi*
13	Desperation_GALC	Desperation	Negative Emotion words	deject*, desolat*, despair*, desperat*, despond*
14	Disappointment_GALC	Disappointment	Negative Emotion words	comedown, disappoint*, discontent*, disenchant*, disgruntl*
15	Disgust_GALC	Disgust	Negative Emotion words	abhor*, avers*, detest*, disgust*, dislik*
16	Dissatisfaction_GALC	Dissatisfaction	Negative Emotion words	dissatisf*, unhapp*
17	Envy_GALC	Envy	Negative Emotion words	envious*, envy*
18	Fear_GALC	Fear	Negative Emotion words	afraid*, aghast*, alarm*, dread*, fear*
19	Guilt_GALC	Guilt	Negative Emotion words	blame*, contriti*, guilt*, remorse*, repent*
20	Hatred_GALC	Hatred	Negative Emotion words	acrimon*, hat*, rancor*
21	Irritation_GALC	Irritation	Negative Emotion words	annoy*, exasperat*, grump*, indign*, irrita*
22	Jealousy_GALC	Jealousy	Negative Emotion words	covetous*, jealous*
23	Sadness_GALC	Sadness	Negative Emotion words	chagrin*, deject*, dole*, gloom*, glum*

https://drive.google.com/file/d/1SUfSYNyuaWT2i4tQkiyr2rxVeqnh3cQe/view

Benefits Using SEANCE for Text Analysis

01
Comprehensive Analysis

02
Dimensionality Reduction

03
Quantitative Insightsn

04
Sentiment Analysis

Example of Text Analysis Results by SEANCE

Text: “The library's new online catalog system is very user-friendly and efficient. The staff is always willing to help, although the website sometimes loads slowly.”

nwords	negative_adjectives_component	positive_adjectives_component	politeness_component
20	-0.30	0.85	0.25

nwords: The total number of words in the text.

negative_adjectives_component: Indicates the presence and intensity of negative adjectives. A negative value suggests the presence of some negative comments such as "loads slowly".

positive_adjectives_component: Positive value indicates the presence of positive adjectives such as "user-friendly," "efficient," and "help".

politeness_component: Moderate indications of politeness or formal language, as seen in the respectful critique "although the website sometimes loads slowly".

What to do with Text Analysis Results?

Identify Common Issues:

- Aggregate negative feedback, count the frequency of entities mentioned in the text.

Compare user sentiment before and after the implementation of new services:

- Statistical comparison of the sentiment element in user feedback in two periods.

Statistic Test

Wilcoxon Rank Sum Test

- Nonparametric: Does not assume the data follow a specific distribution.
- Independent Samples: Compares two groups that are not related.
- Median Comparison: Tests whether the medians of two groups differ.

https://www.isixsigma.com/dictionary/wilcoxon-rank-sum-test/

Example

Suppose a librarian wants to compare the effects of two different services for youth. They collect data user feedbacks about two services, conduct text analysis to identify sentiment.

- Group A (Service 1): 5 social media posts
- Group B (Service 2): 5 social media posts

The positive sentiment intensity for each social media post in the groups is as follows:

- Group A: 0.3, 0.5, 0.7, 0, 0.2
- Group B: 0, 0, 0.1, 0.5, 0.2

Hypotheses:

- **Null Hypothesis (H0):** The positive sentiment for both services is the same.
- **Alternative Hypothesis (H1):** The positive sentiment differs between the two services.

Example

The Wilcoxon Rank Sum Test (Mann-Whitney U test) results are as follows:

- U Statistic: 12.0
- p-value: 0.31473916375196484

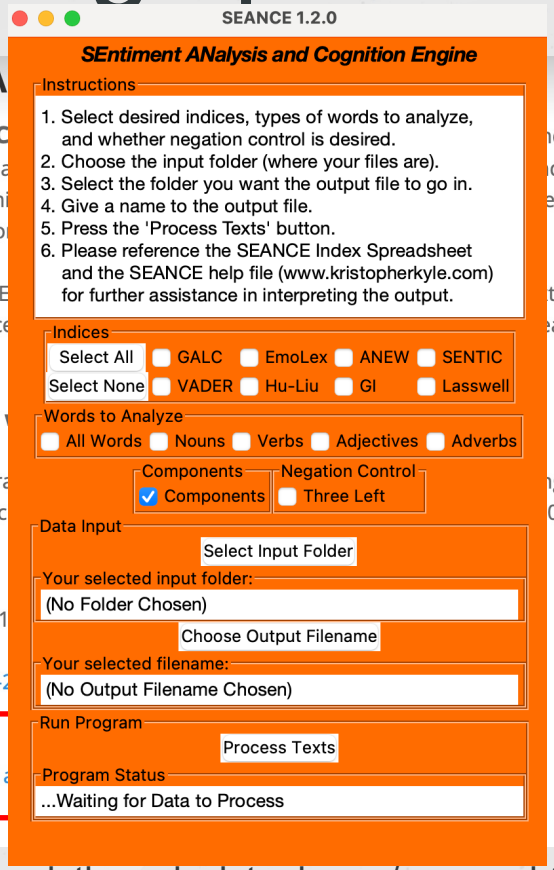
Interpretation:

Given the p-value (0.31473916375196484), which is greater than the common significance level of 0.05, we fail to reject the null hypothesis. This suggests that there is not enough evidence to conclude that the positive sentiment differs between the two services.

02

Demonstrations

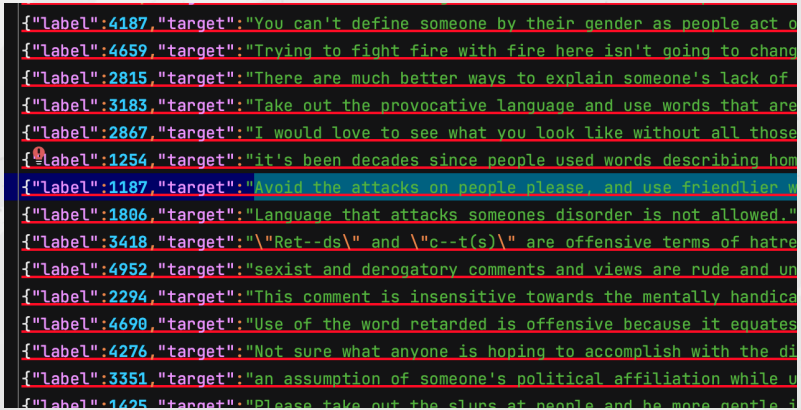
Setting up SEANCE



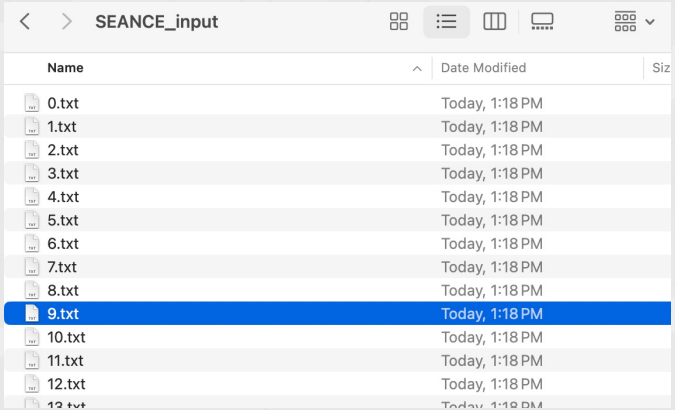
https://www.linguisticanalysistools.org/seance.html

Data Process

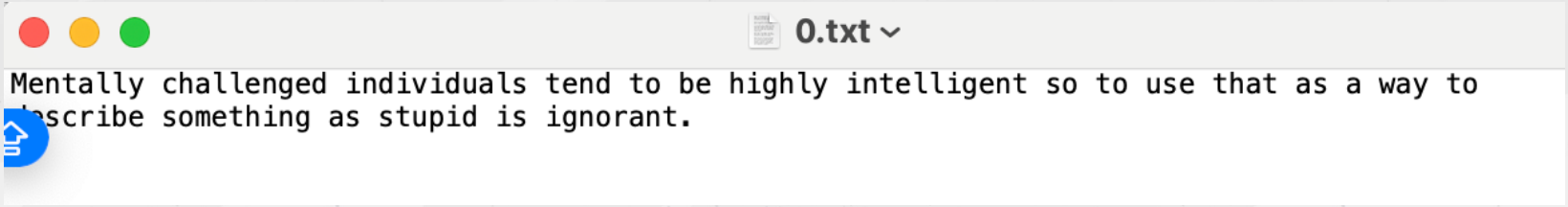
Original json file



Input Folder

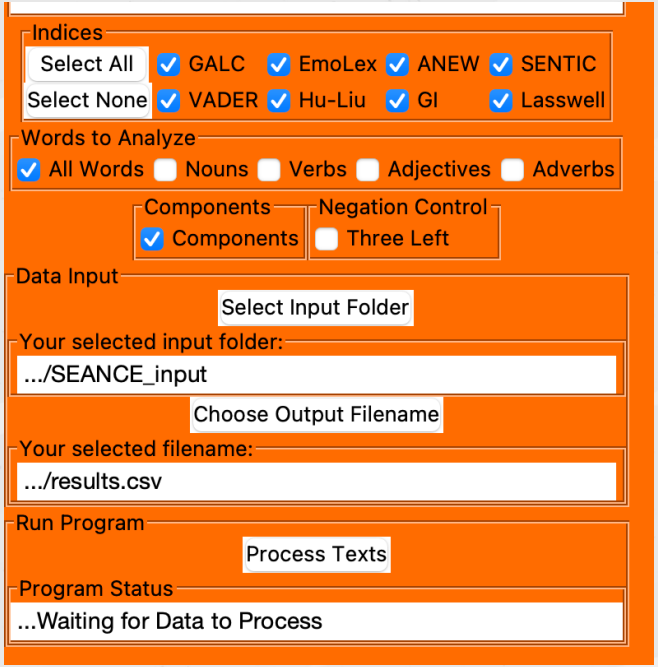


Input text



Run SEANCE

1. Select desired indices, types of words to analyze, and whether negation control is desired.
2. Choose the input folder (where your files are).
3. Select the folder you want the output file to go in.
4. Give a name to the output file.
5. Press the 'Process Texts' button.
6. Please reference the SEANCE Index Spreadsheet and the SEANCE help file (www.kristopherkyle.com) for further assistance in interpreting the output



Results

index	nwords	negative_adjectives_component	social_order_component	action_component	positive_adjectives_component	joy_component	affect_friends_and_family_compo
29	30	-0.3982002	0.2012212121212120	0.3634696969696970	-0.0387374242424242	0.0	0.19785466666
15	21	-1.147	0.2266190476190480	0.8042058823529410	0.9431960784313730	0.0	0.152575630252
14	7	0.2837824	0.1268571428571430	0.0	-0.3924	0.0	
28	11	0.0	0.0886363636363636	1.206090909090910	-0.545	0.0	0.19982909090
16	13	-0.6678144	0.1953846153846150	0.0391538461538461	-0.2345307692307690	0.0358461538461538	0.270243076923
17	9	0.4409794666666670	2.3206666666666700	1.4075555555555600	-0.39349	0.0	0.244235555555
13	21	0.3748611142857140	0.2062492997198880	0.4625252100840340	0.2449661764705880	0.0	0.164339047619
12	15	1.8794587333333300	0.401	0.2891833333333330	0.4408033333333330	3.898953333333330	0.325072000000
10	20	-0.6091144	1.2434833333333300	0.4037333333333330	-0.23105	0.0	0.505
38	20	-0.519	0.9343	0.089	-0.44645	0.0	
39	7	-0.6678144	0.3718571428571430	0.3309285714285710	0.0523535714285713	0.0	
11	17	5.010202505882350	0.2627058823529410	0.4415359477124180	-0.431095	0.0	
76	19	3.3110381263157900	0.5106421052631580	0.6307578947368420	-0.2455539473684210	0.0	0.053888421052
62	15	5.632556666666670	0.5212666666666670	0.0506	-0.30302	0.0	0.258893333333
89	20	-0.6881004000000000	0.3394642857142860	0.1780428571428570	-0.1782000000000000	0.117	0.643
88	15	1.924603866666670	1.6812151515151500	0.7792727272727270	-0.3150009090909090	0.0	0.131982666666
63	20	2.9326564000000000	0.0743	0.1737833333333330	-0.0554299999999999	0.40185	0.373
77	21	1.9244887955182100	0.1094033613445380	0.7764963585434170	-0.472733893557423	0.79804	0.181557142857
49	17	-0.9295284235294120	0.0822941176470588	0.1494264705882350	1.7055964705882400	0.0551176470588235	
61	8	1.0332279333333300	0.193375	0.2225	-0.281765	0.0	
75	22	3.647365268686870	0.0886363636363636	0.7065833333333330	-0.2527677777777780	0.0	0.126970101010
74	5	0.556277	0.0	0.9860666666666670	-0.216725	0.0	0.400960000000

Wilcoxon Rank Sum Test

```python
import pandas as pd
from scipy.stats import wilcoxon
data = pd.read_csv("/Users/xiaoying/Desktop/pythonProject/workshop/results
data_1 = data[data["category"] == 1]
data_1 = data_1.drop(columns=["index", "category","text"])
data_0 = data[data["category"] == 0]
data_0 = data_0.drop(columns=["index", "category","text"])
print(data_1.head(5))
stat, p = wilcoxon(data_1, data_0)
print('Statistics=%.3f, p=%.3f' % (stat, p))
# interpret
alpha = 0.05
if p > alpha:
    print('Same distribution (fail to reject H0)')
else:
    print('Different distribution (reject H0)')
```

Statistics=109.0, p=0.032139802890493 57
Lust_GALC has significant difference
Pleasure/Enjoyment_GALC
Statistics=72.5, p=4.184119148623244e-15
Pleasure/Enjoyment_GALC has significant difference
Pride_GALC
Statistics=0.0, p=0.17971249487899976
Pride_GALC has no significant difference
Relaxation/Serenity_GALC
Statistics=19.0, p=0.6773318478512569
Relaxation/Serenity_GALC has no significant difference
.....

Results Interpretation

	All	Discussion	Identity	Media	Meme	Hobby
Textual factors						
Tokens	↑↑↑	↑↑↑	↑↑↑		↑↑↑	↑↑↑
1st person pronouns	↑↑↑	↑↑↑	↑↑		↑↑	
2nd person pronouns	↑↑↑	↑↑↑	↑↑↑	↑↑↑	↑↑↑	↑↑↑
Communication Form	↑↑↑	↑↑↑	↑↑↑		↑↑↑	↑↑↑
Interjection	↑↑↑	↑↑↑	↑↑↑		↑	↑↑↑
Sentiment factors						
Neutral words	↓↓↓	↓↓↓	↓↓↓		↓	↓↓
Negative words	↑↑↑	↑↑↑	↑↑↑		↑↑	↑↑↑
Positive words	↑↑↑	↑↑↑	↑↑↑	↑	↑	↑↑
Arousal words	↑↑↑	↑↑↑	↑↑↑		↑	↑
Vice words	↑↑↑	↑	↑↑↑			↑↑
Pain words	↑↑↑	↑↑↑	↑↑↑			↑↑↑
Topic factors						
Enlightment words	↑↑↑	↑↑↑	↑↑↑		↑↑↑	↑↑↑
Action words	↑↑↑	↑↑↑	↑↑↑	↑	↑↑↑	
Evaluation words	↑↑↑	↑↑↑	↑↑↑		↑	↑↑↑
Certainty words	↑↑↑	↑↑↑	↑↑↑		↑↑↑	↑↑
Power words	↑↑↑	↑↑↑	↑↑		↑	↑

Q&A

Thank you!