

SEANCE: Sentiment Analysis Program
(By Kristopher Kyle)

<http://www.kristopherkyle.com/seance.html>

Taken from SEANCE 1.04 User Manual

Contents

Quick Start Guide	1
SEANCE Index Overview	2
Source databases	5
General inquirer	5
Lasswell	5
Geneva affect label coder.....	5
Affective norms for English words (ANEW)	5
EmoLex.....	6
SenticNet.....	6
Valence aware dictionary for sentiment reasoning (VADER)	7
Hu-Liu polarity	7
SEANCE component scores	7

Crossley, Scott. A., Kristopher Kyle, and Danielle S. McNamara. 2017. “Sentiment Analysis and Social Cognition Engine (SEANCE): An Automatic Tool for Sentiment, Social Cognition, and Social Order Analysis.” Behavior Research Methods, Vol. 49, No. 3, pp. 803-821.

Quick Start Guide

1. Download SEANCE from www.kristopherkyle.com/seance.html
2. Unzip the SEANCE file.
3. Make sure that all of the texts you want to process are in a single folder and are in plain-text format (SEANCE only processes .txt files).
4. Click on the SEANCE icon to open the program.
5. Choose the indices that you wish to include in your analysis. The default choice includes 20 component scores (see below for greater details of the indices reported by SEANCE).

6. Click the “Select Input Folder” button to select the folder where your text files are located.
7. Click the “Choose Output Filename” button to select a name and location for SEANCE to write its output to.
8. Click the “Process Texts” button to start the textual analysis.
9. Congratulations! Your text files have been turned into numbers. Now the fun begins.

SEANCE Index Overview

Below is a summary of the indices included in SEANCE. For a more detailed treatment of the indices, please see the supplemental SEANCE Index Spreadsheet, which is available at www.kristopherkyle.com/seance.html. This sheet has a list of all the general variables and their categories along with a list of all possible indices reported by SEANCE (taking into account negation and part of speech tags).

SEANCE contains a number of pre-developed word vectors developed to measure sentiment, cognition, and social order. These vectors are taken from freely available source databases such as SenticNet (Cambria, Speer, Havasi, & Hussain, 2010; Cambria, Havasi, & Hussain, 2012) and EmoLex (Mohammad & Turney, 2010, 2013). For many of these vectors, SEANCE also provides a negation feature (i.e., a contextual valence shifter; Polanyi & Zaenen, 2006) that ignores positive terms that are negated. The negation feature, which is based on Hutto and Gilbert (2014), checks for negation words in the 3 words preceding a target word. In SEANCE, any target word that is negated is ignored within the category of interest. For example, if SEANCE processes the sentence *He is not happy* the lexical item *happy* will not be counted as positive emotion word. This method has been shown to identify approximately 90% of negated words (Hutto & Gilbert, 2014). SEANCE also includes the Stanford part of speech (POS) tagger (Toutanova, Klein, Manning, & Singer, 2003) included in Stanford CoreNLP (Manning et al.,

2014). The POS tagger allows for POS tagged specific indices for nouns, verbs, and adjectives. POS tagging is an important component of sentiment analysis because unique aspects of sentiment may reside more strongly adjectives (Hatzivassiloglou & McKeown 1997; Hu & Liu 2004; Taboada, Anthony, & Voll 2006) or verbs and adverbs (Benamara et al. 2007; Sokolova & Lapalme 2008; Subrahmanian & Reforgiato 2008). SEANCE reports on both POS variables and non-POS variables. Many of the vectors in SEANCE, for example, are neutral with regard to POS. This allows for SEANCE to accurately process poorly formatted texts that cannot be accurately analyzed by a POS tagger. We briefly discuss the source databases used in SEANCE below. Table 1 provides an overview of the categories reported in SEANCE and the source databases that report on each category.

Table 1

An overview of the text categories in SEANCE and their source databases

Category	GALC	EmoLex	ANEW	SenticNet	VADER	Hu-Liu	Harvard IV-4	Lasswell	Total
Action							2		2
Arousal	2		2				1		5
Arts and academics							2	1	3
Cognition				2			9	7	18
Communication							4	1	5
Dominance, respect, money, and power			2				4	24	30
Economics, politics, and religion							9	10	19
Effort							10	2	12
Evaluation							9		9
Feeling/emotion				1			2		3
Negative emotion words	18	5			1	2	7	6	39
Other affect					2			4	6
Physical							13	1	14
Positive emotion words	13	2		1	1	2	4	3	26
Quality and quantity							9		9
Reference							4		4
Social Relations	4	1					17	3	25
Surprise	1	2							3
Time and space							13	1	14
Valence/polarity			2	1		1			4
Total	38	10	6	5	4	5	119	63	250

Source databases

General inquirer. SEANCE includes the Harvard IV-4 dictionary lists used by The General Inquirer (GI; Stone et al., 1966). The GI lists are the oldest manually constructed lists still in widespread use and include 119 word lists organized into 17 semantic categories containing over 11,000 words. These categories include 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. The lists were developed for content analysis by social, political, and psychological scientists. Greater detail on the categories and available word lists is available at <http://www.wjh.harvard.edu/~inquirer/homecat.htm>.

Lasswell. SEANCE also includes the Lasswell dictionary lists (Lasswell & Namewirth, 1969; Namewirth & Weber, 1987), which are included in the GI. Included are 63 word lists organized into 9 semantic categories. These categories include power, rectitude, respect, affection, wealth, well-being, enlightenment, and skill. Additional information on these categories and their supporting word lists is available at <http://www.wjh.harvard.edu/~inquirer/homecat.htm>.

Geneva affect label coder. The Geneva Affect Label Coder (GALC) is a database that is comprised of lists of words pertaining to 36 specific emotions and 2 general emotional states (positive and negative; Scherer, 2005). The specific emotion lists include anger, guilt, hatred, hope, joy, and humility. In practice, many of these word lists are quite small and should be used only on larger texts that provide greater linguistic coverage (in order to avoid non-normal distributions of data).

Affective norms for English words (ANEW). The Affective Norms for English Words (ANEW) database (Bradley & Lang, 2009) includes affective norms for valence, pleasure, arousal, and

dominance (Osgood, Suci, & Tanenbaum, 1957). Unlike LIWC and GI word lists, ANEW word lists have associated sentiment scores that are positive if the score is above 5 and negative if below 5 (and neutral if at 5). Bradley and Lang collected norms using the Self-Assessment Manikin system (Lang, 1980) to collect norms for 1,033 English words.

EmoLex. EmoLex (Mohammad & Turney, 2010, 2013) consists of lists of words and bigrams that evoke particular emotions (e.g., anger, anticipation, disgust, fear, joy, sadness, surprise and trust). Additionally, EmoLex include lists of words and bigrams that generally evoke negative and positive emotions. Word and bigram lists were compiled from entries in the *Macquarie Thesaurus* (Bernard, 1986) that were also frequent in the Google n-gram corpus (Brants & Franz, 2006), the WordNet Affect Lexicon (Strapparava & Valitutti, 2004), and the General Inquirer (Stone, Dunphy, Smith & Ogilvie, 1966). Mohammad and Turney then used Amazon Mechanical Turk to determine which emotions (if any) were evoked by each word or bigram. The ten lists each include between 534 (for surprise) and 3,324 (for negative emotions) entries. EmoLex has been used to examine emotions in mail and email (Mohammad & Yang, 2011) and for investigating emotion in fiction writing (Mohammad, 2012).

SenticNet. SenticNet (Cambria, Speer, Havasi, & Hussain, 2010; Cambria, Havasi, & Hussain, 2012) is a database extension of WordNet (Fellbaum, 1998) consisting of norms for around 13,000 words with regard to four emotional dimensions (sensitivity, aptitude, attention and pleasantness) based on work by Plutchik (2001) and norms for polarity. Unlike LIWC, GI, or ANEW, SenticNet scores were calculated using semi-supervised algorithms and the scores are thus not a gold-standard resource. SenticNet was designed to build and improve upon SentiWordNet (Esuli & Sebastiani, 2006) using a number of data-refining techniques.

[Valence aware dictionary for sentiment reasoning \(VADER\)](#). The Valence Aware Dictionary for Sentiment Reasoning (VADER) is a rule-based sentiment analysis system (Hutto & Gilbert, 2014) developed specifically for shorter texts found in social media contexts (e.g., Twitter or Facebook). VADER uses a large list of words and emoticons that include crowd-sourced valence ratings. Additionally, the VADER system includes a number of rules that account for changes in valence strength due to punctuation (i.e., exclamation points), capitalization, degree modifiers (e.g., intensifiers), contrastive conjunctions (i.e., *but*), and negation words that occur within three words before a target word. VADER has been used to accurately classify valence in social media text, movie reviews, product reviews, and newspaper articles (Hutto & Gilbert, 2014).

[Hu-Liu polarity](#). SEANCE includes two large polarity lists compiled by Hu and Liu (2004) for the purposes of sentiment analysis. The Hu-Liu word lists were developed specifically for product reviews and social texts. The positive word list includes 2006 entries, while the negative word list includes 4,783 entries. Both lists were constructed through bootstrapping processes in WordNet. The Hu-Liu lists were used to successfully predict whether product reviews were positive or negative (Hu & Liu, 2004; Liu, Hu, & Cheng, 2005).

[SEANCE component scores](#). One potential pitfall with the SEANCE tool is the sheer number of indices that it reports. With the potential for each index to report results for all words, nouns, verbs, and adjectives in addition to each of these having the potential to be negated, the SEANCE tool can report on almost 3,000 indices. Such a large number of indices for the uninitiated can be unwieldy. Thus, we developed component scores derived from the SEANCE indices to provide users with more manageable options if desired.

To compute the component scores, we adopted an approach similar to Graesser, McNamara, & Kulikowich (2011) and Crossley, Kyle, and McNamara (in press). We conducted a

principle component analysis (PCA) to reduce the number of indices selected from SEANCE to a smaller set of components, each of which was comprised of a set of related features. The PCA, based on the Movie Review corpus, clustered the indices into groups that co-occurred frequently allowing for a large number of variables to be reduced into a smaller set of derived variables (i.e., the components). This gives us two approaches to assessing sentiment. A micro-feature approach (i.e., the indices individually) and a macro-feature approach (i.e., the indices aggregated into components).

For inclusion into a component, we set a conservative cut off for the eigenvalues to ensure that only strongly related indices would be included in the analysis (i.e., .40). For inclusion in the analysis, we first checked that all variables were normally distributed. We then controlled for multicollinearity between variables (defined as $r > .90$) so that selected variables were not measuring the exact construct. After conducting the factor analysis, we set the variance explained for each component at 1% for inclusion into SEANCE. Components that explained less than 1% of the variance were removed. For the included component scores, we used the eigenvalues for each included index to create weighted component scores. In total, we developed 20 component scores. These components explained 56% of the variance in the Movie Review corpus. The 20 components are summarized in Table 2.

Table 2

Description of component scores

Component	Label	Number of indices	Key indices
1	Negative adjectives	18	NRC negative adjectives, NRC disgust adjectives, NRC anger adjectives, GI negative adjectives, Lu Hui negative adjectives
2	Social order	11	RC ethics verbs, GI need verbs, RC rectitude words
3	Action	9	GI ought verbs, GI try verbs, GI travel verbs, GI descriptive action verbs
4	Positive adjectives	9	Lu Hui positive adjectives, Vader positive, GI positive adjectives, Laswell positive affect adjectives
5	Joy	8	NRC joy adjectives, NRC anticipation adjectives, NRC surprise adjectives
6	Affect for friends and family	9	Laswell affect nouns, Laswell participant affect, GI kin noun, GI affiliation nouns
7	Fear and disgust	8	NRC disgust nouns, NRC negative nouns, NRC fear, NRC anger
8	Politics	7	GI politics, GI politics nouns, Laswell power
9	Polarity nouns	7	Polarity nouns, Pleasantness nouns, Aptitude nouns
10	Polarity verbs	4	Polarity verbs, Aptitude verbs, Pleasantness verbs
11	Virtue adverbs	5	Laswell rectitude gain adverbs, GI concerns for hostility adverbs, Laswell sureness adverbs
12	Positive nouns	4	Lu Hui nouns
13	Respect	4	Laswell respect nouns
14	Trust verbs	5	NRC trust verbs, NRC joy verbs, NRC positive verbs
15	Failure	5	Laswell power loss verbs, GI failure verbs
16	Well being	4	Laswell well-being physical nouns, Laswell well-being total
17	Economy	4	GI names adjectives, GI economy adjectives, GI economy all
18	Certainty	6	GI quantity, GI overstatement, Laswell if, Laswell sureness nouns
19	Positive verbs	3	Lu Hui positive verbs
20	Objects	4	GI objects, GALC being touched