Procesamiento del Lenguaje Natural, Revista nº 58, marzo de 2017, pp. 133-140 recibido 01-12-2016 revisado 10-01-2017 aceptado 12-02-2017

Lingmotif: A User-focused Sentiment Analysis Tool

Lingmotif: una Herramienta de Análisis de Sentimiento Enfocada en el Usuario

Antonio Moreno-Ortiz
Universidad de Málaga
29071 Málaga, Spain
amo@uma.es

Abstract: In this paper, we describe Lingmotif, a lexicon-based, linguistically-motivated, user-friendly, GUI-enabled, multi-platform, Sentiment Analysis desktop application. Lingmotif can perform SA on any type of input texts, regardless of their length and topic. The analysis is based on the identification of sentiment-laden words and phrases contained in the application's rich core lexicons, and employs context rules to account for sentiment shifters. It offers easy-to-interpret visual representations of quantitative data, as well as a detailed, qualitative analysis of the text in terms of its sentiment. Lingmotif can also take user-provided plugin lexicons in order to account for domain-specific sentiment expression. As of version 1.0, Lingmotif analyzes English and Spanish texts. Lingmotif thus aims to become a general-purpose Sentiment Analysis tool for discourse analysis, rhetoric, psychology, marketing, the language industries, and others.
Keywords: Sentiment analysis, content analysis, discourse analysis, digital humanities.

Resumen: En este artículo se describe Lingmotif, una aplicación de Análisis de Sentimiento multi-plataforma, con interfaz gráfica de usuario amigable, motivada lingüísticamente y basada en léxico. Lingmotif efectúa Análisis de Sentimiento sobre cualquier tipo de texto, independientemente de su tamaño o tema. El análisis se basa en la identificación en el texto de palabras y frases con carga afectiva, contenidas en los diccionarios de la aplicación, y aplica reglas de contexto para dar cabida a modificadores del sentimiento. Ofrece representaciones gráficas fáciles de interpretar de los datos cuantitativos, así como un análisis detallado del texto. Lingmotif también puede utilizar léxicos del usuario a modo de *plugins*, de tal modo que es posible analizar de forma efectiva la expresión del sentimiento en dominios específicos. La versión 1.0 de Lingmotif está preparada para trabajar con textos en español e inglés. De este modo, se conforma como una herramienta de propósito general en el ámbito del Análisis de Sentimiento para el análisis del discurso, retórica, psicología, marketing, las industrias de la lengua y otras.
Palabras clave: Análisis de sentimiento, análisis de contenido, análisis del discurso, humanidades digitales.

1 Introduction[1]

Sentiment Analysis (SA), along with text analytics in general, has experimented increased attention in the last 15 years, no doubt due to the ever-increasing surge of user-generated content (UGC) on the World Wide Web, a vast body of knowledge that companies and organizations seek to sift, probe, and make sense of. Since text is the form that most of this knowledge is encoded as, it is no surprise that text analytics, or text mining, has become the focus of many research efforts.

Such strong interest has resulted in a vast body of technical knowhow, academic publications, and software. Most available SA software, however, is in the form of either code libraries, usually as part of NLP toolkits for developers, such as NLTK (Loper and Bird, 2002), Stanford CoreNLP (Manning et al., 2014), Apache OpenNLP (Morton et al., 2005), or end-user, “black-box”, commercial applications and services, mostly focused on the

[1] This research was supported by Spain’s MINECO through the funding of project Lingmotif2 (FFI2016-78141-P).

analysis of user-generated content, such as that produced by social networking sites.

These tools usually make use of supervised, Machine Learning techniques, which implies that they either require that users train the classifiers on their own data sets (in the case of developer libraries) or rely on the trained algorithms offered by commercial products.

The disadvantage of the first type of tools is that users are required to possess certain programming skills, whereas the latter offer no indication of what was found in the text that was used to classify a text as positive or negative. Furthermore, such tools are almost invariably geared toward short texts where opinion or sentiment is known to be expressed: user reviews, tweets or other online UGC. Their applicability to longer, multi-topic texts is simply not considered.

However, the automatic identification and analysis of sentiment in texts is interesting not just for sifting online UGC sources, but for many other applications, such as content and discourse analysis. Lingmotif attempts to tackle such needs by taking a radically different approach, and opens a door to a wider range of applications than current Sentiment Analysis tools offer. It can be used as classifier in the “traditional” sense, but it can also be used as a general-purpose text analysis tool that will show the sentiment profile of long texts, identify and clearly display sentiment expressions, provide a number of useful text metrics, compare texts alongside one another, produce analysis of a time series, and more.

Lingmotif is available as desktop application for the Windows, Mac OS, and Linux platforms, and is free for non-commercial uses. Currently, it supports English and Spanish input text, with ongoing development for French, German and Italian.

1.1 Approaches to Sentiment Analysis

Two approaches are distinguished to tackle the automatic analysis of semantic orientation. Most systems, as mentioned above, make use of statistical, Machine Learning techniques, mostly supervised methods, where the SA problem is seen as one of classification: a text is either positive or negative (sometimes finer-grained categories) to be classified under one of these classes. In these systems, a set of tagged examples of the type the classifier is meant to deal with (the training set) is used to train the classifying algorithms. The algorithm is then evaluated against a second set of tagged examples (the evaluation set), and accuracy metrics (in terms of precision and recall) are obtained that allow such systems to be compared in terms of performance. A classic example of such systems is Pang, Lee and Vaithyanathan (2002). Machine Learning classifiers generally work well with the type of content they have been trained for, but their performance drops, almost to chance, when they are used with other types of texts (Taboada et al., 2011). Several approaches have been used to adapt ML-based classifiers to various subject domains (Aue and Gamon, 2005, Choi, Kim and Myaeng, 2009), but the problem remains.

The second approach involves the use of rich lexical sources where sentiment-carrying lexical items are listed. The task of determining the semantic orientation of a text, consists of identifying such items in the input texts, perhaps analyze their context, and perform calculations on the identified items. A classic example of this type of system is Turney (2002).

2 *Sentiment Analysis with Lingmotif*

Lingmotif is a lexicon-based SA system, since it uses a rich set of lexical sources and analyzes context in order to identify sentiment laden text segments and produce two scores that qualify a text from a SA perspective. In a nutshell, it breaks down a text into its constituent sentences, where sentiment-carrying words and phrases are searched for, identified, and assigned a valence (i.e., a sentiment index). The complete analysis process is explained in section 4 below.

2.1 Levels of analysis

Sentiment Analysis, as a classification task, can take different text units as the object of classification. Traditionally, most SA systems have focused on document-level classification, that is, their function is to classify an input text as positive or negative (Turney, 2002, Pang, Lee and Vaithyanathan, 2002). Fewer systems have taken sentences (Wiebe and Riloff, 2005) or clauses (Wilson, Wiebe and Hwa, 2004, Thet et al., 2009) as classification segments. The reason why most systems perform document-level classification is simply that they are designed to classify short documents, traditionally, user reviews. Lingmotif analyzes text at the sentence-level, which, from a linguistic point of view, leaves much to be desired, since many sentiment indicators operate extra-sententially.

2.2 Sentiment shifters

Sentiment shifters, or contextual valence shifters, were first proposed by Polanyi and Zaenen (2006) as a mechanism to account for the modification (or *shift*) of the sentiment of a given lexical unit by means of its surrounding context. Since then, they have been implemented in a number of lexicon-based SA systems: Kennedy and Inkpen, 2006, Moreno Ortiz et al., 2010, Taboada et al., 2011. Sentiment can be altered by context in different ways: it can be intensified, diminished, or it can be inversed altogether. Negation is probably the most relevant shifter (Wiegand et al., 2010), since it usually inverts the polarity of the lexical item it modifies, but intensification and downtoning need also be addressed. In section 3.2 below, we describe Lingmotif's CVS system.

2.3 Analysis modes

Lingmotif uses a simple, but efficient GUI that allows users to select input and options, and launch the analysis (see Figure 1). Results are generated as an HTML document, which is saved to a predefined location and automatically sent to the user's default browser for immediate display.

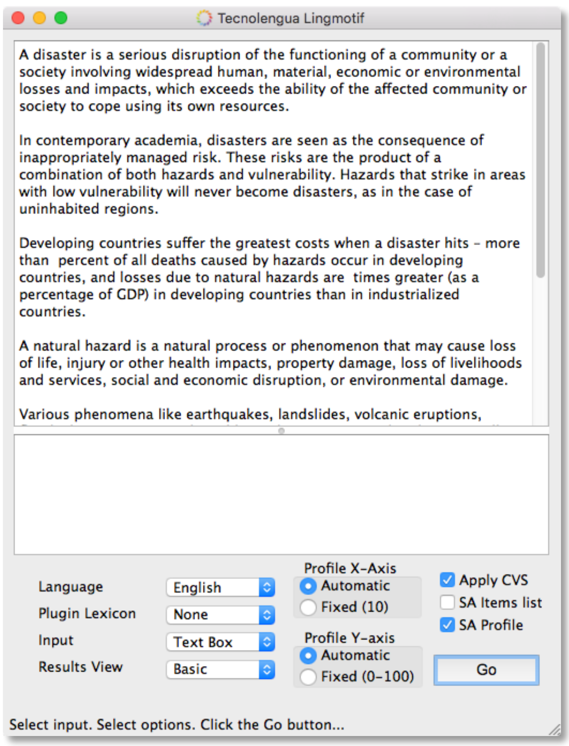

Figure 1: Lingmotif's GUI

Internally, the application generates results as an XML document containing all the relevant data; this XML document is then parsed against one of several available XSL templates, and transformed into the final HTML and Javascript. This interface allows users to simply type or paste a text in the input text area, or load a number of text files to be analyzed. Lingmotif works in either single-document or multi-document mode.

2.3.1 Single-document mode

Whether in single or multi-document mode, Lingmotif will always produce a number of metrics for each individual text, which we list below:

- TSS: Text Sentiment Score: the text's overall sentiment score.
- TSI: Text Sentiment Intensity: the proportion of sentiment vs non-sentiment items. These two are shown graphically (Figure 2)

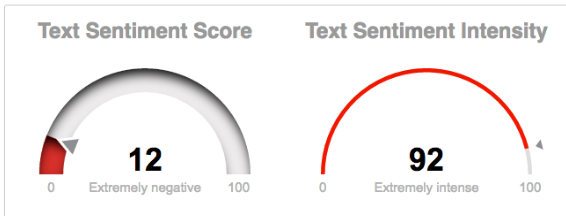

Figure 2: TSS and TSI gauges

- Sentiment Profile: a graphical representation of the text's sentiment "flow". See Figure 3.

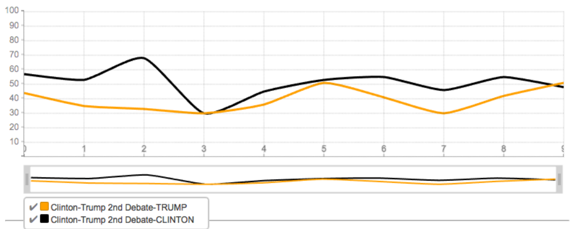

Figure 3: Sentiment profile (parallel mode)

- Text Analysis: several text metrics: number of tokens, types, sentences, lexical and function words, etc.
- Quantitative Sentiment Analysis: a breakdown of the figures that were obtained in order to come up with the TSS and TSI (See Figure 4).
- Detailed Sentiment Analysis: a display of the input text where sentiment items are color coded according to their polarity and

specific data for each are (optionally) displayed (See Figure 5).

Text Analysis

Text Stats			Sentences	Words by Function		Words by Form	
Tokens	Types	T/T Ratio		Lexical	Grammatical	Single Words	Multiwords
6898	1293	18.74%	637	2916	3982	6409	489

Sentiment Analysis

TSS	TSI	CVS segments	Positive items	Negative items	Neutral items	Positive Score	Negative Score
38	88	95	169	298	2416	379	693

Figure 4: Quantitative data

Detailed Sentiment Analysis

Well, I actually agree_with that.

I agree_with everything she said.

I began this campaign because I was so_tired_of seeing such foolish things happen to our country.

This is_a_great country.

This is_a_great land.

I have gotten_to_know the people of the country over the last year-and-a-half that I have been doing this a s a politician.

Figure 5: Detailed sentiment analysis

The most relevant results are the TSS and TSI. These two scores qualify a text in terms of its orientation (TSS) and intensity (TSI). Both are displayed by means of visual, animated gauges at the top of the results page. These gauges also include a category for each, from "extremely negative" to "extremely positive", which makes numeric results readily interpretable by the user (see Figure 2). We describe how these scores are obtained in section 4.

However, for long texts, the Sentiment Profile is a powerful that can provide a quick insight into the text's internal structure and organization in terms of sentiment expression. This graph is interactive: hovering the data points will display the lexical items that make up that particular text segment. The quantitative data tables are also quite useful when comparing texts (see next section), since it readily offers useful common text metrics, such as type/token ratio.

2.3.2 Multi-document analysis

Multi-document mode is enabled simply by loading multiple files or one multi-document file (i.e., a file where each line is assumed to be a –short– document, such as tweets or user reviews).

When in multi-document mode, Lingmotif will analyze documents one by one, generating one HTML file for each, although they will not be displayed on the browser, just saved to the output folder. When the analysis is finished, a single results page will be displayed. This page is a summary of results, and is different from the single-document results page: the gauges for TSS and TSI are now the average for the analyzed set and the detailed analysis section contains a quantitative analysis of each of the files in the set. The first column in this table shows the title of the document (file name without extension) as a hyperlink to the HTML file for that particular file.

Multi-document mode has several modes of operation:

- Classifier (default): a stacked bar graph and data table are offered showing classification results based on their TSS category. The graph offers a visualization of results; both its legend and the graph itself are interactive (see Figure 6).

TOTAL DOCUMENTS								
469								
TSS CLASSIFICATION								
Extremely	Very	Fairly	Slightly	Neutral	Slightly	Fairly	Very	Extremely
29	40	53	48	207	28	32	14	18
THREE-WAY CLASSIFICATION								
Negative				Neutral	Positive			
170 (36.25%)				207 (44.14%)	92 (19.62%)			
BINARY CLASSIFICATION								
Negative					Positive			
274 (58.42%)					195 (41.58%)			

Figure 6: Classifier data table

- Series: the set of loaded files is assumed to be in order, chronological (time series) or otherwise. Each data point in the Sentiment Analysis Profile represents one document. The data point is the average TSS for that particular document.
- Parallel: produces a graph with one line for each file (this mode is limited to 15 documents). This is useful to compare sentiment flow in texts side by side.
- Merge: this option merges all loaded individual files in one single text.

3 *Lexical resources*

Lingmotif's performance is directly proportionally to the quality of its lexical resources. The creation of our current lexical resources has been our focus for many years. Work on Spanish the lexicon started with the Sentitext project (Moreno-Ortiz et al., 2010) and was further expanded, refined adapted to Lingmotif's format during the Lingmotif 1 project, along with the creation of the English resources.

For each language, Lingmotif requires the following resources:

- A full-coverage core sentiment lexicon which contains both single words and

multiword expressions.
- A set of context rules, where sentiment shifters are defined using a template approach.
- A part of speech tagger.
- A lemmatizer, used for generating inflected forms during lexicon imports or updates.
- Optionally, a plugin lexicon can be used to account for domain-specific sentiment expression.

3.1 Core lexicon

The core lexicon is the most important resource. A lexical item in a Lingmotif lexicon (whether a single word or a multiword expression) is defined by a specification of its form, part of speech, and valence. The valence is an integer from -5 to -2 for negatives and 5 to 2 for positives. The item's form can either be a literal string or a lemma (represented between angled brackets). As for the part-of-speech specification, Lingmotif uses the Penn Treebank tag set for all languages. A wildcard (`ALL`) can be used for cases where all possible parts of speech for that lemma share the same valence. Figure 7 below shows some examples.

```
in_cold_blood,RB,-4
<kill>_time,VB,0
<kill>,ALL,-3
broke,JJ,-2
```

Figure 7: Lingmotif lexicon format

The creation process of Lingmotif's core lexicons basically involves merging freely available sentiment lexicons, adapt the merged list to our format and refine it using corpus analysis techniques and sheer heuristics. For English, we merged items from The Harvard General Inquirer (Stone and Hunt, 1963), MPQA (Wilson, Wiebe and Hoffmann, 2005), and Bing Liu's Opinion Lexicon (Hu and Liu, 2004). These resources were expanded by using a thesaurus and derivational generation rules. These resources, however, are characterized by their lack of attention to multiword expressions. We then used a number non SA-specific lexical resources, including common idioms from Wiktionary, which we tagged manually for valence. Ultimately, Lingmotif's lexicons are the result of intensive lexicographical work. A similar processed was followed for the Spanish language. The English lexicon contains over 77,000 entries (word forms) and nearly 500 context rules. The Spanish lexicon contains 207,000 word forms and over 300 context rules.

Sentiment disambiguation is currently dealt with using exclusively formal features: part-of speech tags and multi-word-expressions. MWEs usually include words that may or may not have the same polarity of the expression. including such expressions can solve disambiguation for many cases. For example, we can classify as negative the word "kill" and then include phrases such as "kill time" with a neutral valence. When this is not possible, the options are to include it with the more statistically probable polarity or simply leave it out when the chances of getting the item with one polarity or another are similar.

3.2 Context rules

Context rules are Lingmotif's mechanism to deal with sentiment shifters. They work by specifying words or phrases that can appear in the immediate vicinity of the identified sentiment word. Basically, we use the same approach as Polanyi and Zaenen (2006) or Kennedy and Inkpenn (2006), or Taboada (2011): we use simple addition or subtraction (of integers on a -5 to 5 scale in our case).

In Lingmotif, every rule specifies the following:
- The part of speech and polarity of the sentiment word.
- The form, location (left or right), and span (in number of words) of the shifter.
- The result of the rule application.

Currently, Lingmotif uses over 400 such rules for each language. Table 1 below shows examples of all types of sentiment shifters according to the effect they produce on the resulting text segment.

Shift type	Example Context Rule
Inversion	`NN,-,avoid*,LR,5,INV0`
	`JJ,+-,not,L,2,INV0`
Intensification	`JJ,+-,seriously,L,2,INT3`
	`VB,+-,may_well,L,1,INT1`
Downtoning	`NN,-,mild,L,2,DOW1`
	`NN,+-,a_bit,L,2,DOW1`

Table 1: Context rules types and examples

Lingmotif's context rules were compiled by extensive corpus analysis, studying concordances of common polarity words (adjectives, verbs, nouns, and adverbs), and then testing the rules against texts to further improve and refine them.

When a context rule is matched, the resulting text segment is marked as a single unit and assigned the calculated valence, as specified in the rule. Multiple context rule matching is possible and not handled at present: as soon as a rule is matched, no further rules are searched and the rule is applied. It would definitely be interesting to improve this by establishing a priority system for rules.

3.3 Plugin lexicons

As many researchers have pointed out (Aue and Gamon, 2005, Turney 2002, Read, 2005, Pang and Lee, 2008), sentiment is very often dependent on topic, or domain. Attributes such as size, weight, or location, for example, can be regarded as positive or negative, or neither, depending on whether we are discussing electronic gadgets, hotels or movies.

Being a general-purpose SA system, Lingmotif provides a flexible mechanism to adapt to specific domains by means of user-provided lexicons. Lexical information contained in plugin lexicons overrides Lingmotif's core lexicon. When a plugin lexicon is selected for analysis, the plugin lexicon is searched first. If a word or phrase is found there, the core lexicon will not be searched for that item, and its information in the plugin lexicon will be used. Thus, plugin lexicons can be used to provide domain-specific sentiment items, but also to override polarity assignment in the core lexicon, for whatever reasons.

Plugin lexicons have exactly the same format as the core lexicon. In order to import a plugin lexicon, it must first be created as a UTF-8 encoded CSV file, which is then imported. Updating a plugin lexicon simply involves modifying the source CSV file and importing it again. Any number of plugin lexicons can be created in Lingmotif, but only one can be used for a given analysis.

4 *Analysis process*

As mentioned above, Lingmotif's results are obtained by identifying sentiment-laden words and multiword expressions, analyzing their contexts for sentiment shifters, and weighing sentiment against non-sentiment items. In this section, we describe this process in detail.

The analysis process is the following:

1. Preprocessing: text is scanned for common abbreviations, contractions and misspelling.
2. Tokenization: both sentence-level and word-level.
3. Multiword identification: n-grams are matched against the list multiword expressions contained in the plugin lexicon (if selected) and core lexicon. Identified MWEs are marked and assigned their valence.
4. Polarity words identification: individual words are looked up in the lexicons and assigned their valence if found.
5. Context rule matching: identified polarity words and MWEs are matched against the list of context rules. If a rule matches the word's context, the whole text segment is marked and tagged as a unit of type CVS (context valence shifter). This process is repeated twice, in order to account for cumulative sequences of shifters, such as "very very good".
6. TSS and TSI calculation and category assignment.
7. Generation of internal XML document, which contains the results data for the input text.
8. Generation of Javascript code for the graphical components.
9. Generation of the final results HTML document.

Calculation of the Text Sentiment Intensity (TSI) and Text Sentiment Score (TSS) metrics deservers further discussion. Valences found in lexical units are added and weighed against the number of neutral *lexical* units. Therefore, we do not use the simpler "term-counting method" (Kennedy and Inkpen, 2006): a particularly intense unit can have more weight in the overall score than two less intense units. Also, function words do not enter into the equation. Lingmotif offers all the figures employed to come up with the final scores in the "Text Analysis" section (see Figure 4 above).

First TSI is calculated as the proportion of sentiment vs non-sentiment items (or rather, their added scores). TSI calculation takes text length into account in order to capture the fact that the ratio of sentiment to non-sentiment items is necessarily lower the longer the text. A max_tsi factor is created which places stronger weight in sentiment words in longer texts. TSI is then calculated as

$$TSI = \frac{pos_{score}\% + neg_{score}\%}{\max \_tsi}$$

TSS is then calculated as

$$TSS = \frac{\sum \begin{cases} v * sw, if\ v \neq 50 \\ v, otherwise \end{cases}}{v_n * 100}$$

where v is the value of each lexical unit, and sw is the sentiment weight, a factor inversely proportional to the max_tsi previously calculated.

5 Performance evaluation

Even though Lingmotif was not conceived as a classifier, it can compete with ML-based classifiers in terms of performance. In this section, we employ some readily available SA-tagged data sets to evaluate Lingmotif's performance, obtaining outstanding results. Table 2 below shows the confusion matrix and precision, recall and f-measure figures summarizing the evaluation results against the STS-Gold data set (Saif et al., 2013), which is a data set specifically designed to serve as a gold standard in Sentiment Analysis of Twitter text.

	POS	NEG	TOTAL
POS	522	97	619
NEG	254	1125	1379
	Evaluation		
	Precision	Recall	F-measure
	0.92	0.82	0.87

Table 2: STS-Gold data set results

It must be mentioned that Lingmotif's is not a binary classifier, therefore a number of the documents were actually classified as binary. Specifically, 294 of the 1,379 negative documents and 145 of 719 positive documents. Neutral documents are simply coerced randomly as positive or negative in order to be able to compare results alongside binary classifiers.

Results were similar with other Twitter data sets, such as UMICH SI650[2] or the Stanford Twitter Sentiment Test Set (STS-Test) (Go, Bhayani, and Huang, 2009), which we show in Table 3 below.

	POS	NEG	TOTAL
POS	139	22	161
NEG	37	132	169
	Evaluation		
	Precision	Recall	F-measure
	0.86	0.78	0.82

Table 3: STS-Test data set results

6 Conclusions

Being an end-user tool, evaluating Lingmotif requires more than accuracy figures exclusively. Aspects such as usability and adequacy to specific tasks should also be discussed. We believe that the application also addresses such aspects successfully. All in all, Lingmotif is a platform that offers many possibilities for the analysis of texts from a Sentiment Analysis perspective. Its lexicon-based approach, coupled with a careful curation of its resources results in highly accurate results.

Even so, there are many ways in which it can be improved. Specifically, sentiment disambiguation is only partially dealt with. Current context rules are limited in their expressive power, and would no doubt benefit if semantic categories could be specified, rather than simply words or lemmas. In general, deeper semantic analysis of context would be necessary to improve on current results, both at the sentence level and at the text level.

7 References

Aue, A., and M. Gamon. 2005. Customizing Sentiment Classifiers to New Domains: A Case Study. In *Recent Advances in Natural Language Processing (RANLP-05).* Borovets, Bulgaria.

Choi, Y., Y. Kim, and M. Sung-Hyon. 2009. Domain-Specific Sentiment Analysis Using Contextual Feature Generation. In *Proceeding of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion*, 37–44. Hong Kong, China: ACM.

Go, A., R. Bhayani, and L. Huang. 2009. Twitter Sentiment Classification Using Distant Supervision. CS224N Project Report. Stanford.

Hu, M., and B. Liu. 2004. Mining and Summarizing Customer Reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–77. Seattle, WA, USA: ACM.

Kennedy, A., and D. Inkpen. 2006. Sentiment Classification of Movie Reviews Using Contextual Valence Shifters. *Computational Intelligence*, 22 (2): 110–25.

[2] https://inclass.kaggle.com/c/si650winter11/data

Loper, E., and S. Bird. 2002. NLTK: The Natural Language Toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, 1, 63–70. ETMTNLP '02. Stroudsburg, PA, USA: ACL.

Manning, C., M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 55–60. ACL.

Moreno-Ortiz, A., Á. Pérez Pozo, and S. Torres Sánchez. 2010. Sentitext: Sistema de Análisis de Sentimiento para el Español. *Procesamiento de Lenguaje Natural,* 45: 297–98.

Morton, T., G. Bierner, J. Kottmann, and J. Baldridge. 2005. OpenNLP: A Java-Based NLP Toolkit. Available at https://opennlp.apache.org.

Pang, B., and L. Lee. 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval,* 2 (1–2): 1–135.

Pang, B., L. Lee, and S. Vaithyanathan. 2002. Thumbs up?: Sentiment Classification Using Machine Learning Techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, 10, 79–86. ACL.

Polanyi, L., and A. Zaenen. 2006. Contextual Valence Shifters. In *Computing Attitude and Affect in Text: Theory and Applications, Shanahan*, James G., Qu, Y., and Wiebe, J., 20:1–10. The Information Retrieval Series. Dordrecht, The Netherlands: Springer.

Read, J. 2005. Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentiment Classification. In *Proceedings of the ACL Student Research Workshop*, 43–48. ACLstudent '05. Stroudsburg, PA, USA: ACL.

Saif, H., M. Fernández, Y. He, and H. Alani. 2013. Evaluation Datasets for Twitter Sentiment Analysis: A Survey and a New Dataset, the STS-Gold. Turin, Italy.

Stone, P. J., and E. B. Hunt. 1963. A Computer Approach to Content Analysis: Studies Using the General Inquirer System. In *Proceedings of the May 21-23, 1963, Spring Joint Computer Conference*, 241–56. AFIPS '63 (Spring). New York, NY, USA: ACM.

Taboada, M., J. Brooks, M. Tofiloski, K. Voll, and M. Stede. 2011. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2): 267–307.

Thet, T. T., J. Na, C. S. G. Khoo, and S. Shakthikumar. 2009. Sentiment analysis of movie reviews on discussion boards using a linguistic approach. In *Proceeding of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion*. 81–84. Hong Kong, China: ACM.

Turney, P. D. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting of the ACL*, 417–24. Philadelphia, USA.

Wiebe, J., and E. Riloff. 2005. Creating Subjective and Objective Sentence Classifiers from Unannotated Texts. In *Computational Linguistics and Intelligent Text Processing*, 486–97. Lecture Notes in Computer Science 3406. Springer Berlin Heidelberg.

Wiegand, M., A. Balahur, B. Roth, D. Klakow, and A. Montoyo. 2010. A Survey on the Role of Negation in Sentiment Analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, 60–68. NeSp-NLP '10. Stroudsburg, PA, USA: ACL.

Wilson, T., J. Wiebe, and P. Hoffmann. 2009. Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. *Computational Linguistics*, 35(3): 399–433.

Wilson, T., J. Wiebe, and R. Hwa. 2004. Just How Mad Are You? Finding Strong and Weak Opinion Clauses. In *Proceedings of the 19th National Conference on Artifical Intelligence*, 761–67. AAAI'04. San Jose, California: AAAI Press.