Home	Research	Publications	Lexicons	Labeled Data

Invited Talks	Contact

Sentiment and Emotion Lexicons

This page lists various word association lexicons that capture word-sentiment, word-emotion, and word-colour associations. They can be used for analysing emotions in text. See Terms of Use at the bottom of the page.

Contact: Saif M. Mohammad (saif.mohammad@nrc-cnrc.gc.ca)

Manually Created Lexicons

These lexicons are created by manual annotation. The lexicons with real-valued scores are created using Best-Worst Scaling, producing fine-grained, yet highly reliable annotation values.

Large Manually Created Emotion and Sentiment Lexicons

Lexicon	Version	# of Terms	Categories	Association Scores	Method of Creation
1a. NRC Word-Emotion Association Lexicon (also called NRC Emotion lexicon or EmoLex). README. Explore the interactive visualization. Homepage of the Lexicon. Also available in over 40 other languages here. The sense-level annotations provided by individual annotators for the eight emotions can also be obtained.					
	0.92 (2010)	14,182 unigrams (words)	**sentiments:** negative, positive	0 (not associated) or 1 (associated)	**Manual:** By crowdsourcing **Domain:** General
		~25,000 senses	**emotions:** anger, anticipation, disgust, fear, joy, sadness, surprise, trust	not associated, weakly, moderately, or strongly associated	

Papers:

Crowdsourcing a Word-Emotion Association Lexicon, Saif Mohammad and Peter Turney, *Computational Intelligence,* 29 (3), 436-465, 2013.   Paper (pdf)   BibTeX

Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon, Saif Mohammad and Peter Turney, In *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text,* June 2010, LA, California.   Paper (pdf)   BibTeX   Presentation

1b. NRC Emotion Intensity Lexicon (aka Affect Intensity Lexicon), created using Best-Worst Scaling.
The NRC Emotion Intensity Lexicon is a list of English words and their associations with four basic emotions (anger, fear, sadness, joy). Lexicon homepage.

Word Affect Intensities. Saif M. Mohammad. In *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, May 2018, Miyazaki, Japan.
Paper (pdf)   BibTeX   Presentation

2. NRC Valence, Arousal, Dominance Lexicon, created using Best-Worst Scaling.
The NRC Valence, Arousal, Dominance Lexicon is a list of English words and their valence, arousal, and dominance scores. Lexicon homepage.

	1 (2018)	~20,000 terms	Valence (positive--negative) Arousal (excited--calm) Dominance (powerful--weak)	0 (lowest V/A/D) to 1 (highest V/A/D)	**Manual:** By crowdsourcing **Domain:** General

Paper:

Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. Saif M. Mohammad. In *Proceedings of* the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, July 2018.
Paper (pdf)   BibTeX

Manually Created Sentiment Composition Lexicons

These lexicons include sentiment scores for two- and three-word expressions as well as scores for their constituent words.

Lexicon	Version	# of Terms	Categories	Association Scores	Method of Creation

1. Sentiment Composition Lexicon of Negators, Modals, and Adverbs (SCL-NMA), aka SemEval-2016 General English Sentiment Modifiers Lexicon, created using Best-Worst Scaling (aka MaxDiff)

	1.0 (Feb. 2016)	~3200 terms	**sentiments:** negative, positive	Real-valued score between -1 (most negative) to	**Manual.** By crowdsourcing and using Best-Worst Scaling.

				1 (most positive)	**Domain:** General

Papers:

- The Effect of Negators, Modals, and Degree Adverbs on Sentiment Composition. Svetlana Kiritchenko and Saif M. Mohammad, In *Proceedings of the NAACL 2016 Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media (WASSA),* June 2014, San Diego, California.
  Paper (pdf)   BibTeX   Presentation

- Capturing Reliable Fine-Grained Sentiment Associations by Crowdsourcing and Best-Worst Scaling. Svetlana Kiritchenko and Saif M. Mohammad. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* June 2016. San Diego, CA.
  Paper (pdf)   BibTeX   Presentation

- Semeval-2016 Task 7: Determining Sentiment Intensity of English and Arabic Phrases. Svetlana Kiritchenko, Saif M. Mohammad, and Mohammad Salameh. In *Proceedings of the International Workshop on Semantic Evaluation (SemEval '16).* June 2016. San Diego, California.
  Paper (pdf)   BibTeX   Presentation   Task Website

2. SemEval-2015 English Twitter Sentiment Lexicon, created using Best-Worst Scaling (aka MaxDiff)

	1.0 (Feb. 2015)	~1500 terms	**sentiments:** negative, positive	Real-valued score between -1 (most negative) to 1 (most positive)	**Manual.** By crowdsourcing and using Best-Worst Scaling. **Domain:** Twitter

Paper:

- SemEval-2015 Task 10: Sentiment Analysis in Twitter. Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif M Mohammad, Alan Ritter, and Veselin Stoyanov. In Proceedings of the ninth international workshop on Semantic Evaluation Exercises (SemEval-2015), June 2015, Denver, Colorado.
  Paper (pdf)   BibTeX

- Sentiment Analysis of Short Informal Texts. Svetlana Kiritchenko, Xiaodan Zhu and Saif Mohammad. *Journal of Artificial Intelligence Research,* volume 50, pages 723-762, August 2014.
  Paper (pdf)   BibTeX

  This data was used in SemEval-2015 Task 10 (Sentiment Analysis in Twitter), subtask E - Determining strength of association of Twitter terms with positive sentiment (or, degree of prior polarity). Task description, trial data, test data, and other details available here.

3. Sentiment Composition Lexicon of Opposing Polarity Phrases (SCL-OPP) aka SemEval-2016 English Twitter Mixed Polarity Lexicon, created using Best-Worst Scaling (aka MaxDiff)

	1.0 (Feb. 2016)	~1200 terms	**sentiments:** negative, positive	Real-valued score between -1 (most negative) to 1 (most positive)	**Manual.** By crowdsourcing and using Best-Worst Scaling. **Domain:** Twitter

Paper:

- Sentiment Composition of Words with Opposing Polarities. Svetlana Kiritchenko and Saif M. Mohammad. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* June 2016. San Diego, CA.
  Paper (pdf)   BibTeX   Poster

- *Happy Accident:* A Sentiment Composition Lexicon for Opposing Polarities Phrases. Svetlana Kiritchenko and Saif M. Mohammad. In *Proceedings of the 10th edition of the Language Resources and Evaluation Conference*, May 2016, Portorož (Slovenia).
  Paper (pdf)   BibTeX   Poster

- Semeval-2016 Task 7: Determining Sentiment Intensity of English and Arabic Phrases. Svetlana Kiritchenko, Saif M. Mohammad, and Mohammad Salameh. In *Proceedings of the International Workshop on Semantic Evaluation (SemEval '16)*. June 2016. San Diego, California.
  Paper (pdf)   BibTeX   Presentation   Task Website

Large Manually Created Word-Colour Association Lexicon

Lexicon	Version	# of Terms	Categories	Association Scores	Method of Creation

1. NRC Word-Colour Association Lexicon

	0.92 (2011)	~14,000 words	**colours:** black, ☐ blue, ☐ brown, ☐ green, ☐ grey, ☐ orange ☐ purple, ☐ pink, ☐ red, white, yellow	0 (not associated) or 1 (associated)	**Manual:** Crowdsourcing on Mechanical Turk. **Domain:** General
		~25,000 senses		not, weakly, moderately, or strongly associated	

Papers:

Colourful Language: Measuring Word-Colour Associations, Saif Mohammad, In Proceedings of the ACL 2011 Workshop on Cognitive Modeling and Computational Linguistics (CMCL), June 2011, Portland, OR.   Paper (pdf)   BibTeX   Presentation

Even the Abstract have Colour: Consensus in Word-Colour Associations, Saif Mohammad, In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, June 2011, Portland, OR.   Paper (pdf)   BibTeX   Poster

Automatically Created Lexicons

These lexicons are automatically extracted from large amounts of text using co-occurrence information. For example, the Hashtag Emotion Lexicon is generated from tweets and the score for a word--emotion pair is a quantification of the word's tendency to co-occur with the emotion-word hashtag. These are usually much larger than manually created lexicons. They have higher coverage, especially of terms often seen in the corpus that the lexicon is extracted from. However, the emotion scores can be less accurate than those in the manually created lexicons above.

Large Automatically Generated Word-Emotion Association Lexicon

Lexicon	Version	# of Terms	Categories	Association Scores	Method of Creation
1. NRC Hashtag Emotion Lexicon. The Hashtag Emotion Corpus (aka Twitter Emotion Corpus, or TEC) used to create the lexicon.					
	0.2 (2013)	16,862 unigrams (words)	**emotions:** anger, anticipation, disgust, fear, joy, sadness, surprise, trust	Real-valued score between 0 (not associated) to ∞ (maximally associated)	**Automatic:** From tweets with emotion word hashtags. **Domain:** Twitter

Papers:

Using Hashtags to Capture Fine Emotion Categories from Tweets. Saif M. Mohammad, Svetlana Kiritchenko, Computational Intelligence, Volume 31, Issue 2, Pages 301-326, May 2015. Paper (pdf)   BibTeX

#Emotional Tweets, Saif Mohammad, In Proceedings of the First Joint Conference on Lexical and Computational Semantics (*Sem), June 2012, Montreal, Canada.   Paper (pdf)   BibTeX

Large Automatically Generated Word-Sentiment Association Lexicons

Lexicon	Version	# of Terms	Categories	Association Scores	Method of Creation
1. NRC Twitter Sentiment Lexicons (NRC Hashtag Sentiment Lexicons and Sentiment140 Lexicons)					
a. NRC Hashtag Sentiment Lexicon					
	1.0 (2013)	54,129 unigrams 316,531 bigrams 308,808 pairs	**sentiments:** negative, positive	Real-valued score between -∞ (most negative) to ∞ (most positive)	**Automatic**: From tweets with sentiment word hashtags. **Domain:** Twitter
b. NRC Hashtag Affirmative Context Sentiment Lexicon and NRC Hashtag Negated Context Sentiment Lexicon					
	1.0 (2014)	Affirmative contexts: 36,357 unigrams Negated contexts: 7,592 unigrams Affirmative contexts: 159,479 bigrams Negated contexts: 23,875 bigrams	**sentiments:** negative, positive	Real-valued score between -∞ (most negative) to ∞ (most positive)	**Automatic:** From tweets with sentiment word hashtags. Separate entries for affirmative and negated contexts. **Domain:** Twitter
c. Emoticon Lexicon aka Sentiment140 Lexicon (note that this is sentiment lexicon drawn from emoticons, and is not an emotion lexicon)					
	1.0 (2014)	62,468 unigrams 677,698 bigrams 480,010 pairs	**sentiments:** negative, positive	Real-valued score between -∞ (most negative) to ∞ (most positive)	**Automatic**: From tweets with emoticons. **Domain:** Twitter
d. Sentiment140 Affirmative Context Lexicon and Sentiment140 Negated Context Lexicon					

| | 1.0 (2014) | Affirmative contexts: 45,255 unigrams Negated contexts: 9,891 unigrams | **sentiments:** negative, positive | Real-valued score between -∞ (most negative) to ∞ (most positive) | **Automatic:** From tweets with sentiment word hashtags. Separate entries for affirmative and negated contexts. **Domain:** Twitter |
| | | Affirmative contexts: 240,076 bigrams Negated contexts: 34,093 bigrams | | | |

Papers (describing the four NRC Twitter Lexicons listed above):

Sentiment Analysis of Short Informal Texts. Svetlana Kiritchenko, Xiaodan Zhu and Saif Mohammad. *Journal of Artificial Intelligence Research,* volume 50, pages 723-762, August 2014.
Paper (pdf)   BibTeX

NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets, Saif M. Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu, In Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013), June 2013, Atlanta, USA.
Paper (pdf)   BibTeX   System Description and Downloads   Poster   Slides

NRC-Canada-2014: Recent Improvements in Sentiment Analysis of Tweets, Xiaodan Zhu, Svetlana Kiritchenko, and Saif M. Mohammad. In Proceedings of the eigth international workshop on Semantic Evaluation Exercises (SemEval-2014), August 2014, Dublin, Ireland.
Paper (pdf)   BibTeX

These lexicons were used to generate winning submissions for the sentiment analysis shared tasks of SemEval-2013 Task 2 and SemEval-2014 Task 9.

2. Yelp and Amazon Sentiment Lexicons

   a. Yelp Restaurant Sentiment Lexicon
      (created from the Yelp Dataset -- from the subset of entries pertaining to these restaurant-related businesses)

| | 1.0 (2014) | 39,274 entries for unigrams (includes affirmative and negated context entries) | **sentiments:** negative, positive | Real-valued score between -∞ (most negative) to ∞ (most positive) | **Automatic:** From customer reviews on Yelp.com. **Domain:** Restaurant |

		276,651 entries for bigrams			

The Yelp Word–Aspect Association Lexicons are also made available.

b. Amazon Laptop Sentiment Lexicon

1.0 (2014)	26,577 entries for unigrams (includes affirmative and negated context entries)	**sentiments:** negative, positive	Real-valued score between -∞ (most negative) to ∞ (most positive)	**Automatic:** From customer reviews on Amazon.com. **Domain:** Laptop
	155,167 entries for bigrams			

Paper (describing the Yelp and Amazon Lexicons):

NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews, Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif M. Mohammad. In Proceedings of the eigth international workshop on Semantic Evaluation Exercises (SemEval-2014), August 2014, Dublin, Ireland.　Paper (pdf)　BibTeX

These lexicons were used to generate winning submissions for the sentiment analysis shared task of SemEval-2014 Task 4.

3. Macquarie Semantic Orientation Lexicon

0.1 (2009)	76,400 terms	**sentiments:** negative, positive	binary distinction: negative or positive	**Automatic:** Using the structure of a thesaurus and affixes. **Domain:** General

Paper:

Generating High-Coverage Semantic Orientation Lexicons From Overtly Marked Words and a Thesaurus, Saif Mohammad, Bonnie Dorr, and Cody Dunne, In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2009),* August 2009, Singapore.　Paper (pdf)　BibTeX　Presentation

Links to commonly accessed resources:

- Homepage of the NRC Word-Emotion Association Lexicon, also called EmoLex.
- Homepage for Sentiment Composition Lexicons.
- Homepage for Arabic sentiment lexicons and corpora.
- Homepage for Best-Worst Scaling (aka MaxDiff) software and annotations
- Homepage for the NRC-Canada sentiment anaysis system.
- The AffectiveTweets Package for the Weka machine learning workbench provides a collection of filters for extracting state-of-the-art features (inlcuding features drawn from lexicons provided here) for sentiment classification/regression and other related tasks.
- *See this poster for an overview of the kind of work I have done in the last few years in Computational Affect.*
- *If you are a student interested in working with me, go here.*

Terms of use:

- The lexicons listed here are available free for research purposes. Cite the papers associated with the lexicons in your research papers and articles that make use of them. (The papers associated with each lexicon are listed below, and also in the READMEs for individual lexicons.)

- If interested in commercial use of any of these lexicons, send email to Saif M. Mohammad (Senior Research Officer at NRC and creator of these lexicons): saif.mohammad@nrc-cnrc.gc.ca and Pierre Charron (Client Relationship Leader at NRC): Pierre.Charron@nrc-cnrc.gc.ca. A nominal one-time licensing fee may apply.

- In news articles and online posts on work using these lexicons, cite the appropriate lexicons. For example: "This application/product/tool makes use of the <resource name>, created by <author(s)> at the National Research Council Canada." (The creators of each lexicon are listed below. Also, if you send us an email, we will be thrilled to know about how you have used the lexicon.) If possible hyperlink to this page: http://saifmohammad.com/WebPages/lexicons.html

- If you use a lexicon in a product or application, then acknowledge this in the 'About' page and other relevant documentation of the application by stating the name of the resource, the authors, and NRC. For example: "This application/product/tool makes use of the <resource name>, created by <author(s)> at the National Research Council Canada." (The creators of each lexicon are listed below. Also, if you send us an email, we will be thrilled to know about how you have used the lexicon.) If possible hyperlink to this page: http://saifmohammad.com/WebPages/lexicons.html

- Do not redistribute the data. Direct interested parties to this page: http://saifmohammad.com/WebPages/lexicons.html

- National Research Council Canada (NRC) disclaims any responsibility for the use of the lexicons listed here and does not provide technical support. However, the contact listed above will be happy to respond to queries and clarifications.