**The science behind the Tone Analyzer Service**

The Tone Analyzer service is based on the theory of psycholinguistics, a field of research that explores the relationship between linguistic behaviors and psychological theories

Many people naturally read a message and judge the tone conveyed by the sender. The question is: Can a computer detect the tones conveyed in a message accurately and automatically? This is one of the many challenging questions to which researchers in the artificial intelligence and cognitive sciences fields are seeking answers.

## Overview of related research

Research has shown a strong and statistically significant correlation between word choices and personality, emotions, attitudes, intrinsic needs, values, and thought processes. Several researchers found that people vary in how often they use certain categories of words when writing for blogs, essays, and tweets and that these communication mediums can help predict aspects of personality.Most of these prior works are based on the **Linguistic Inquiry and Word Count (LIWC) psycholinguistics** **dictionary**[Tausczik & Pennebaker, 2010](http://www.ibm.com/watson/developercloud/doc/tone-analyzer/science.shtml#bib-tausczik), and [Pennebaker et al., 2007](http://www.ibm.com/watson/developercloud/doc/tone-analyzer/science.shtml#bib-pennebaker). The LIWC is used to find psychologically meaningful word categories from word usage in writing. This research serves as the basis for IBM's work on the Tone Analyzer service.

## Approach to tone analysis

The following list describes the service's approach to computing a scorecard of language tones

* **Emotional tone** to derive emotion scores from text, we use a stacked generalization-based ensemble framework. Stacked generalization is a general method of using a high-level model to combine lower-level models to achieve greater predictive accuracy. Features such as n-grams (unigrams, bigrams and trigrams), punctuation, emoticons, curse words, greeting words (such as hello, hi, and thanks), and sentiment polarity are fed into state-of-the machine learning algorithms to classify emotion categories.
* **Social tone** consists of the Big Five personality characteristics of openness, agreeableness, and conscientiousness.
* **Language tone** is calculated from the linguistic analysis based on learned features.

# Understand your tone score

|  |  |  |  |
| --- | --- | --- | --- |
| **Emotion** | **Description** | **Low value** | **High Value** |
| joy | Joy or happiness has shades of enjoyment, satisfaction and pleasure. There is a sense of well-being, inner peace, love, safety and contentment. | Less than 0.5 - less likely to be perceived as joyful. | More than 0.75 - Highly likely to be perceived as joyful. |
| fear | A response to impending danger. It is a survival mechanism that is a reaction to some negative stimulus. It may be a mild caution or an extreme phobia. | Less than 0.5 - less likely to be perceived as scared. | More than 0.75 - Highly likely to be perceived as scared. |
| sadness | Indicates a feeling of loss and disadvantage. When a person can be observed to be quiet, less energetic and withdrawn, it may be inferred that sadness exists. | Less than 0.5 - less likely to be perceived as sad. | More than 0.75 - Highly likely to be perceived as sad. |
| disgust | An emotional response of revulsion to something considered offensive or unpleasant. It is a sensation that refers to something revolting. | Less than 0.5 - less likely to be perceived as disgusted. | More than 0.75 - Highly likely to be perceived as disgusted. |
| anger | Evoked due to injustice, conflict, humiliation, negligence or betrayal. If anger is active, the individual attacks the target, verbally or physically. If anger is passive, the person silently sulks and feels tension and hostility. | Less than 0.5 - less likely to be perceived as angry. | More than 0.75 - Highly likely to be perceived as angry. |

## Emotional tone

Emotional tone is inferred from different types of emotions and feelings that people express in their language. For each of these emotions, the service outputs a score that lies between 0 to 1 that indicates the probability that the emotion came across in the text. In this table, the low value describes the lowest a score a sentence or document can receive and can still be considered to contain that emotion. The high value describes when the sentence or document has a high probability of belonging to that emotion.

## Social tone

Social tone measures the social tendencies in people's writing. Tone Analyzer responds with analysis on five different social tones: openness, conscientiousness, extraversion, agreeableness, and emotional range (or neuroticism). These five social tones are adopted from the Big-five personality model. For each of these categories, the service outputs a score from 0 to 1 that indicates tendency toward the listed behaviors.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tone** | **Description** | **Low value** | **High value** |
| openness | The extent a person is open to experience a variety of activities. | Less than 0.25 - more likely to be perceived as no-nonsense, straightforward, blunt, or preferring tradition and the obvious over the complex, ambiguous, and subtle. | More than 0.75 - more likely to be perceived as intellectual, curious, emotionally-aware, imaginative, willing to try new things, appreciating beauty, or open to change. |
| conscientiousness | The tendency to act in an organized or thoughtful way. | Less than 0.25 - more likely to be perceived as spontaneous, laid-back, reckless, unmethodical, remiss, or disorganized. | More than 0.75 - more likely to be perceived as disciplined, dutiful, achievement-striving, confident, driven, or organized. |
| extraversion | The tendency to seek stimulation in the company of others. | Less than 0.25 - more likely to be perceived as independent, timid, introverted, restrained, boring, or dreary. | More than 0.75 - more likely to be perceived as engaging, seeking attention, needy, assertive, outgoing, sociable, cheerful, excitement-seeking, or busy. |
| agreeableness | The tendency to be compassionate and cooperative towards others. | Less than 0.25 - more likely to be perceived as selfish, uncaring, uncooperative, self-interested, confrontational, skeptical, or arrogant. | More than 0.75 - more likely to be perceived as caring, sympathetic, cooperative, compromising, trustworthy, or humble. |
| emotional range | The extent a personâ€™s emotion is sensitive to the environment. | Less than 0.25 - more likely to be perceived as calm, bland, content, relaxed, unconcerned, or careful. | More than 0.75 - more likely to be perceived as concerned, frustrated, angry, passionate, upset, stressed, insecure, or impulsive. |

.

## Language tone

Describes perceived writing style using these categories: analytical style, reasoning style, and confidence. For each of these categories, the service outputs a score from 0 to 1 that indicates tendency toward being perceived as described in the table.

|  |  |  |  |
| --- | --- | --- | --- |
| **Style** | **Description** | **Low value** | **High value** |
| Analytic | A person's reasoning and analytical attitude about things. | Less than 0.25 - the text contains little or no evidence of analytical tone. | More than 0.75 - more likely to be perceived as intellectual, rational, systematic, emotionless, or impersonal. |
| Confidence | A persons degree of certainty. | Less than 0.25 - the text contains little or no evidence of confidence in tone. | More than 0.75 - more likely to be perceived as assured, collected, hopeful, or egotistical. |
| Tentative | A persons degree of inhibition. | Less than 0.25 - the text contains little or no evidence of tentativeness in tone. | More than 0.75 - more likely to be perceived as questionable, doubtful, limited, or debatable. |