

If You're Happy and You Show It :) Detecting Emotion in Text-Based Communication

Taylor Scott^{1,2}, Ona Anicello^{1,2}, Michael Brooks^{1,2}, Katie Kuksenok^{2,3}, Daniel Perry^{1,2}, John Robinson², Cecilia Aragon^{1,2}

University of Washington HCDE¹, SCCL², CSE³

Purpose

Recent research has indicated the importance of affect in motivating creativity and collaboration; however, the automated detection of emotion in text-based communication is an open problem. This poster explores methods of detecting emotion within the chat log of the Nearby Supernova Factory (SNfactory), an international collaboration of astrophysicists studying supernovae and the expansion of the universe. We are in the process of analyzing this data through manual coding and machine learning techniques, in order to detect and classify emotional content and relate it to events occurring in the group's history.



University of Hawaii 2.2m telescope on the summit of Mauna Kea, Hawaii, U.S. Taken from: *Context-Linked Virtual Assistants for Distributed Teams: An Astrophysics Case Study* (Poon et al., CSCW 2008)

Supernova 1994D in the outskirts of the galaxy NGC 4526. This example of a type 1A supernova shows that at peak brightness they rival the cores of galaxies in luminosity (Hubble Space Telescope photo).

Background

Previous research suggests that the presence of certain types of affect in group collaboration may lead to increased creative thinking and perhaps higher rates of project success. However, relatively little is known about the effectiveness of transmitting affective information in computer-mediated communication (CMC) such as email, chats and discussion boards. Typically, most CMC research has investigated communication of cognitive information. Understanding how emotion is mediated across different time zones and work cultures is especially important as scientific collaborations become more geographically dispersed. This study examines emotional content found in logs from online chat discussions recorded over a five-year time span (2004 to 2008). The chats came from members of an international astrophysics group in the United States and France. Members of the group chatted online (synchronously) to study and measure supernova activity via a remotely-controlled, high-end telescope.

Methods

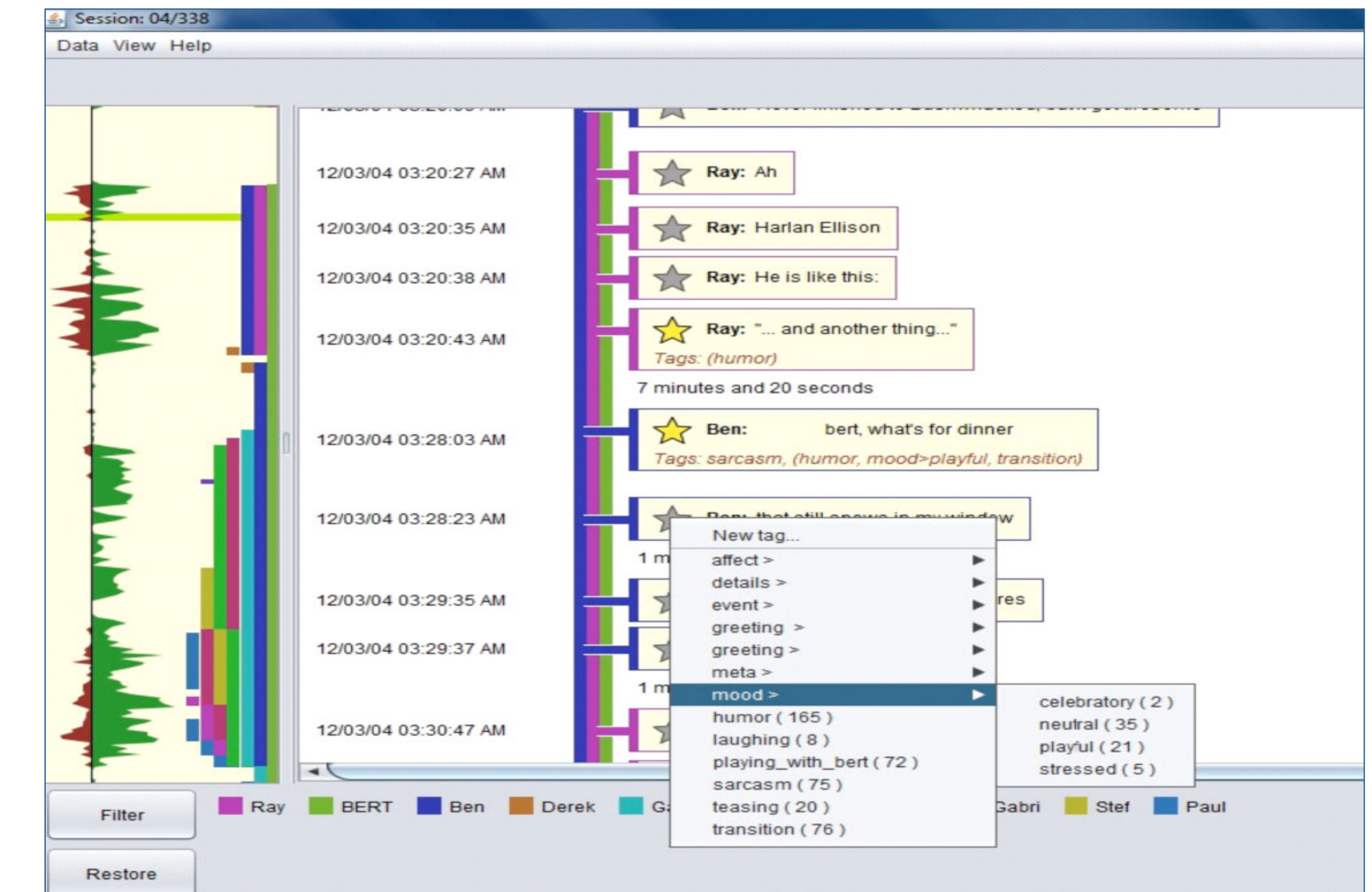
The chat logs are in the process of being coded and analyzed in several ways:

Manual Coding of Data Logs

The chat logs are being coded line-by-line for emotional content utilizing the ETC-Viz application (an in-house software application for coding chat logs pictured to the right). Each line of the chat log was coded with specific affect(s), changes in mood, and event transitions. The affect codes were developed through a combination of grounded theory and a review of the literature on emotion taxonomies.

Automated Processing of Chat Logs and Computer Learning

Once the data is manually coded, it can be processed using Weka (a data mining toolkit) to automate the coding of the entire chat log repository and to correlate emotional content to specific events, such as the discovery of a new supernova or the collaborative solution of a technical, computational, or scientific problem.



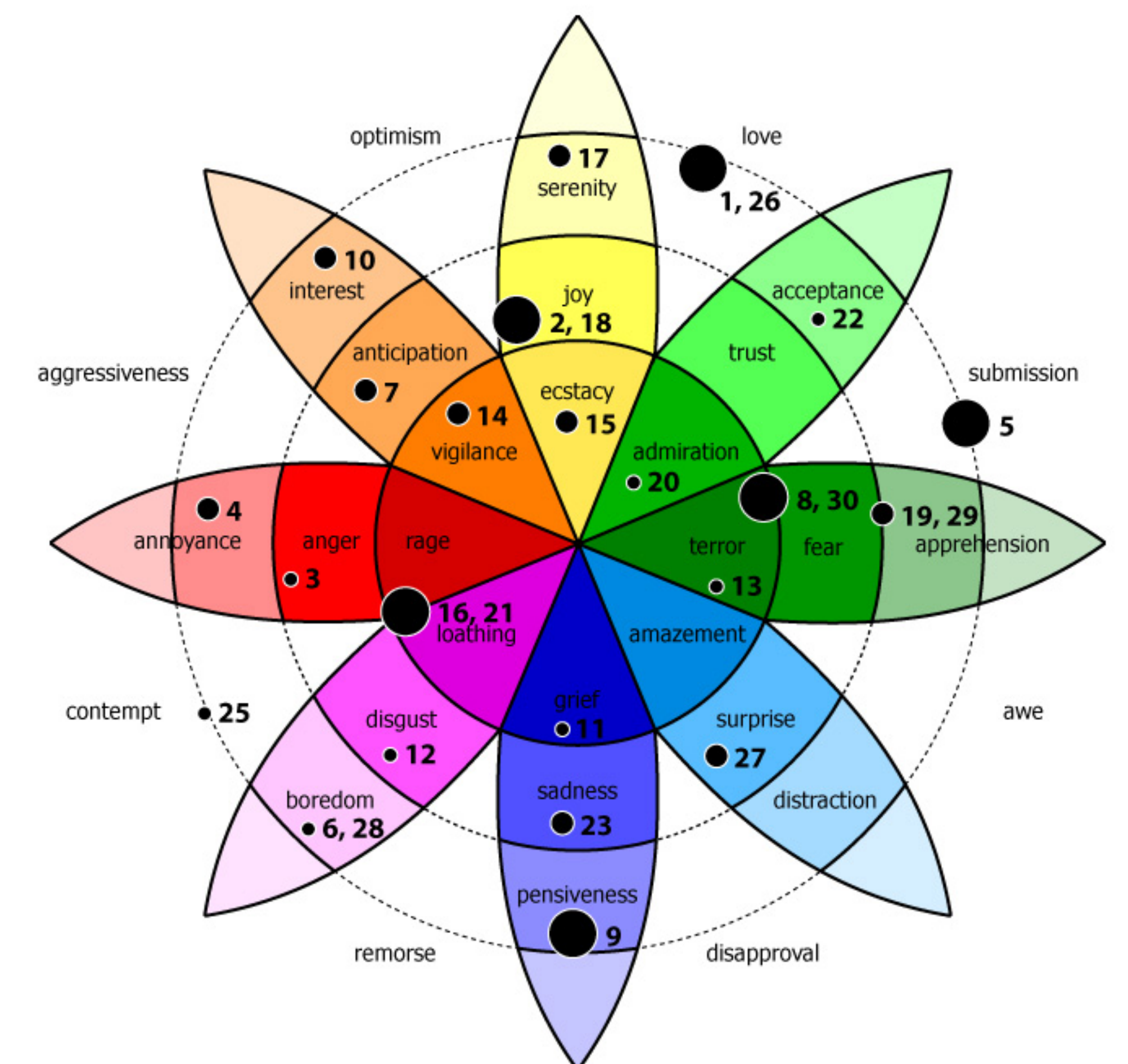
A screenshot from ETC-Viz, the in-house application developed to visually display the chat logs in a timeline view so that the affect codes can be easily applied to the messages by the members of our research group. These tags were then stored on a central server, allowing multiple users to code the same logs simultaneously. In addition to the application of the affect codes, important events, transitions in global mood, and other points of interest were also coded for later analysis.

Results

We are in the process of conducting more extensive data analyses of our coded work. We have developed a coding scheme which has been manually applied to 1300+ messages identified to contain emotion, spanning six nights of chat logs. These coded logs will then be used as a training set for an automated classifier to detect which messages contain emotion. Provided below is a list of the emotion codes being applied to these chat logs, and to the right is a data visualization of those same codes mapped against Robert Plutchick's Wheel of Emotion, a well known emotional coding structure.

1. Agreement	7. Concern	13. Embarrassment	19. Impatient	25. Suggestive
2. Amused	8. Confusion	14. Emphatic	20. Impressed	26. Supportive
3. Anger	9. Considering	15. Excitement	21. Irritation	27. Surprise
4. Annoyed	10. Curious	16. Frustration	22. Relief	28. Tired
5. Apologetic	11. Disappointment	17. Gratitude	23. Sadness	29. Uncertainty
6. Boredom	12. Disgust	18. Happiness	24. Skeptical	30. Worry

The emotion codes contained in this list were derived from existing taxonomies of emotion widely used in psychology, and then augmented based on the contents of our own data set, and to better fit the types of emotional content used during text-based chat. Note that the lines of chat logs found to not contain emotional affect were tagged "no affect".



A visualization of the affect codes applied to Plutchick's Wheel of Emotion. The size of the circles represent the number of times each emotion occurred in the manually coded logs, with each number corresponding to one of the emotions from the list on the left.

For more information, contact Taylor Scott at omni@uw.edu or Cecilia Aragon at aragon@uw.edu or visit <http://depts.washington.edu/sccl>