# VIBES: Visualizing Changing Emotional States in Personal Stories

### April M. Wensel
Department of Computer Science
Pomona College
185 East Sixth Street
Claremont, CA 91711
april.wensel@pomona.edu

### Sara O. Sood
Department of Computer Science
Pomona College
185 East Sixth Street
Claremont, CA 91711
sara@cs.pomona.edu

## ABSTRACT

Online journals (blogs) provide not only an outlet for emotional self-expression, but also a space for social interaction and commiseration through the exchange of personal stories. However, the massive extent of the blogosphere can overwhelm users, restricting their ability to make meaningful connections to fellow bloggers. In this article, we present a system, VIBES, that extracts the important topics from a blog, measures the emotions associated with those topics, and generates a suite of visualizations of this information. Unlike previous research, which has focused on extracting global trends in opinion across the blogosphere, VIBES focuses on depicting the emotional trajectories of the storylines that persist throughout the life experiences of the individual. In user tests, a majority of participants agreed that the visualizations revealed the author's current emotional state and emotional development over time. VIBES has potential applications both in connecting users via shared emotional profiles on social networks, as well as facilitating self-reflection through private user status displays. It also offers a fresh perspective for studying emotions and modeling how they change over time, which has a number of applications in affective computing, including the creation of emotionally responsive interfaces.

## Categories and Subject Descriptors

H.3.5 [**Online Information Services**]: Web-based services; H.3.3 [**Information Search and Retrieval**]: Information filtering

## General Terms

Design, Human Factors

## Keywords

Narrative, Affective Computing, Weblogs, Visualization, Summarization

## 1. INTRODUCTION

The ability to perceive, express, and manage emotions is an important aspect of being human. Psychologists have termed this ability "emotional intelligence," and it has recently become the focus of studies in the fields of both human psychology as well as artificial intelligence.

Two key components of emotional intelligence - according to Daniel Goleman, one of the pioneering authors in the field - are self-awareness and social awareness. Respectively, these relate to the understanding of one's own emotions and the emotions of others [4]. Many believe that these skills are essential to success in life.

The growth of the internet has provided new tools for refining these skills. The blog, which is essentially a public online diary, is one such medium. The author of a blog often writes about personal and emotional experiences, while the reader interprets and responds to these experiences. According to one psychologist, blogs provide the means for "the blogger to mobilize emotional support and reinforce a social identity" [27]. For the blogger, therefore, the expression of emotion can be cathartic and may evoke empathy in the form of reader responses.

Blog readers, on the other hand, may derive comfort from identifying with another's life experiences and may also improve their ability to understand and respond to emotional processes. This can be especially beneficial when reading the blogs of those individuals in one's social network. Both parties, therefore, stand to develop emotional intelligence through participation in the blogosphere. A substantial number of people are in a position to take advantage of these opportunties, with 39% of adult internet users reading blogs and 12% writing in a blog [25]. The number of blog authors is greater among online teenagers, with 28% keeping a blog [24].

Accessing the rich emotional content of blogs, however, is not a simple process. For the reader, the sheer volume of blog entries makes navigating the blogosphere unwieldy. The problem is similar for writers. While they may benefit from expressing their emotions, it is difficult for them to step back and assess their emotional patterns from the whole set of their entries. There is an apparent need for a system that can harvest emotional data from a blog and present it to the people that may care about it.

The goal of this work is the creation of such a system, VIBES[1], that harnesses the power of emotional expression in blogs in order to facilitate both self-reflection and the

---

[1] Visualizations of Individual Bloggers' Emotional States

exchange of emotional intelligence among people. By extracting the topics and their accompanying storylines from a blog, measuring their emotional content automatically, and presenting the results to the user in a useful and meaningful way, VIBES can help users overcome information overload and access specific emotional trajectories found in the life stories of others. Rather than seeking to represent the *vox populi* as other systems have, VIBES will preserve and amplify the voice of the individual.

VIBES generates visualizations that depict the emotional content from an individual blog in three different ways. These visualizations can be placed in a public user profile or displayed privately to the blogger. The existence of an audience for these visualizations is suggested by the popularity of online personality quizzes; people seem to have an affinity for tools that reveal information to them about themselves.

In addition, the visualizations offer a tool with which researchers can model "emotional state transition" [31]. Knowledge of these emotional patterns can assist in building emotionally intelligent interfaces that sense and respond to emotional fluctuations in the user.

## 2. RELATED WORK

There have been a number of systems that analyze changing sentiment across the blogosphere. These systems tend to search for trends in popular opinion about such topics as politicians, consumer products, and current events. One such system is BlogPulse [22]. Some of these global trend trackers only analyze the number of references to a given topic, while others also seek to analyze the positive or negative orientation of the reference. *MoodViews* measures global mood levels across the LiveJournal user base and generates graphs to describe them. One of the *MoodViews* formats is *Moodteller*, which uses natural language processing and machine learning to predict moods based on language used and gets impressive accuracy [20].

There have also been systems that attempt to bring out the underlying emotional commonalities in diverse and otherwise unrelated bloggers. One such project is *We Feel Fine*, which groups blog excerpts that follow a form of the word "feel" with the same emotion [6]. A similar project is *The Dumpster*, which does a similar text search for descriptions of romantic turmoil [3]. The surrounding text is analyzed in order to group break-ups with similar characteristics. In both *We Feel Fine* and *The Dumpster*, there is an artistic interactive interface to navigate through the excerpts. These systems provide an overview of large groups of bloggers as well as access to individual voices.

Most of these relevant past studies look for global trends across large numbers of bloggers, while VIBES focuses on changes concerning topics specific to each individual.

One similar system, Transient Life, provides a summarization of information about an individual's general state of life, which includes an emotional dimension [28]. However, this system requires the user to explicitly talk about their thoughts and moods, while VIBES automatically gleans this data from something the user has already written, i.e. the blog entries.

## 3. THE SYSTEM: VIBES

We propose a system that combines and extends existing technologies in order to visualize the sentiments expressed by an individual blogger about specific topics over time. The system, VIBES, is made up of four primary components: an entry parser, a topic identifier, an emotion classifier, and a set of visualizations (Figure 1). There are three visualization modes: *EmoGraph*, *EmoMeters*, and *EmoCloud*, each of which displays emotion and topic data in a different way. All of the visualization modes rely on the same basic underlying data collection processes, but they differ in exactly which data they use and the ways that they display this data.

The basic system works as follows (Figure 1). Given a LiveJournal username, the system retrieves all of the corresponding user's blog entries and identifies the topics from each entry. Then, the emotion classifier generates sentiment scores for the topics in the context of each entry. It also generates a list of emotion bearing words that occur near the topic. The visualization component processes this information, finds the frequently recurring topics, and generates three visualizations that summarize the emotions found in the storylines corresponding to these topics.

In the sections that follow, the components are described in further detail.

### 3.1 Entry Parser

The first task for the system is to retrieve the entry text from the user's blog. A major problem with processing data from the Internet it is found in the form of HTML, which is designed to enable optimal viewing by humans rather than computers, so it lacks a standard structure. Most blogging platforms provide a partial solution to this problem by supplying RSS feeds of blog posts. RSS feeds are easier for a machine to parse because the data is labeled with standard field names such as <pubDate> and <title>. However, these feeds tend to be limited to only the most recent 25 or so blog entries. Since a primary goal of VIBES is to illustrate change over time, it is important that the system has access to all entries. Therefore, our entry aggregator cannot take advantage of the RSS feeds, but must rather extract the entries directly from the HTML. This requires a certain level of knowledge about the formatting of the entry pages. This format varies so widely across the blogosphere, so we simplify the process at this initial stage by limiting our system's retrieval abilities to one blogging site. This makes parsing the input more manageable.

We chose the blogging site LiveJournal [15] to deploy and test our system for a number of reasons. First of all, LiveJournal is one of the most popular blogging sites, with over 15 million total users [17]. Also, LiveJournal is purportedly designed to be used primarily for personal diary-style journals as opposed to more objective journals focused on current events or practical advice [15]. Finally, LiveJournal has a fairly consistent format for displaying entries. Most LiveJournal navigation and blog entry pages adhere to certain formatting conventions. For example, dates usually precede titles, which precede the entry text. This consistency facilitates the automatic extraction of blog entry text.

The entry extractor automatically retrieves entries for a specified blogger by constructing appropriate URLs and extracting only the blog text from the HTML. Specifically, the
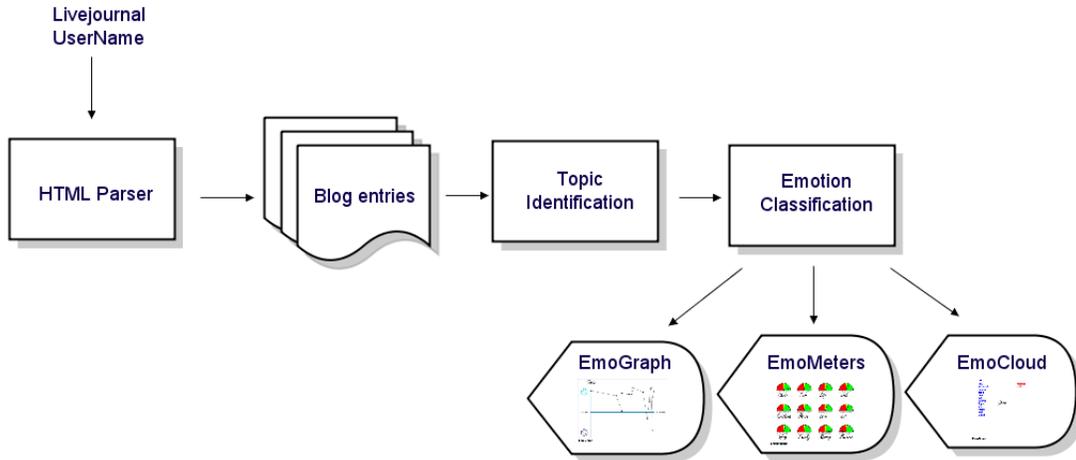
**Figure 1: High-level system architecture.**

system first visits the user's profile page to extract the creation date of the journal and the time of the last post. It then constructs URLs for each month the blog author has maintained the blog. An example URL would be http://username.livejournal.com/2000/03/. Each of these pages displays the times and subjects of each post, with permanent links to each entry.

The system follows each of these links in succession, storing the subject from the link text. On each entry page, the parser reads in lines until it encounters this subject text. It then retrieves the text following the title and preceding the comments section, which is usually signified by a line containing the text "comments." The parser then removes any HTML tags and stores the result as the blog entry text.

## 3.2 Topic Identification

The next step entails identifying the topics in each of the blog entries.

Identifying topics from text is a classic problem in information retrieval. Two basic approaches are topic classification and keyword extraction. The first requires advance knowledge of potential topics [5, 12, 18], while the second does not [1, 19, 9]. VIBES poses a unique challenge because traditional topic identification methods could not easily be applied. The set of candidate topics for each blogger—or even the blogosphere as a whole—is not obvious. We therefore chose the keyword approach.

Specifically, at the heart of the topic identifier is a system called *Balie*, which identifies named entities in a novel semi-supervised way [21]. *Balie*, based on a small list of samples of each type of named entity, uses web searches to generate longer lists of entities of each type. Then it uses the result of this training to identify entities in a given unseen text. Unfortunately, while *Balie* has fairly impressive recall, there are some problems with its precision. The entities it identifies are sometimes words that do not correspond to reasonable topics. For example, it sometimes returns punctuation marks and words like "the," "in," and "or." We did

not have the option to train *Balie* to rectify these issues, as it would require access to the original copyrighted training corpus. Therefore, some further processing steps are required to extract the most descriptive topics to improve precision.
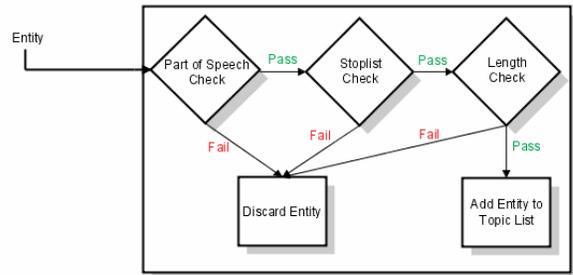


**Figure 2: Individual entity filter.**

The VIBES topic identifier takes several steps to filter these results for the words most likely to represent issues important to the blogger. Given the initial list of entities generated by *Balie*, the topic identifier sends each candidate entity through a filter (Figure 2). The filter first checks to make sure it is not a member of a stoplist, which consists of common words such as "the" and other inappropriate words that often show up as topics, such as "aka." It also checks to make sure that the entity consists of only one word. Although this may seem like an odd decision given that entities can often contain multiple words, there are several motivations for this choice. One problem is that *Balie* sometimes identifies as entities long strings of text that are surrounded by HTML tags. These appear commonly in blogs because people often post images as well as tables representing the results of online personality quizzes the user has taken.

As another discriminating step, the system only accepts topics in the form of nouns or gerunds. The decision to filter

entities for nouns is a common practice in similar systems [26, 32]. Nouns are more appropriate for describing topics. Finally, the system removes topics that are mentioned in fewer than three entries, since they are probably not important and two data points would not make for a revealing visualization of change.

Figure 3 displays a blog entry excerpt. Table 1 illustrates the results of the topic filtering process.

| Topics | | | |
|---|---|---|---|
| . | able | can | **chris** |
| **credit** | **cristina** | **day** | **dr** |
| easier | **eye** | **gardner** | good |
| guess | **idea** | in | **jobs** |
| **life** | little | **lot** | **loyola** |
| made | **may** | **me** | nice |
| now | **office** | one | **place** |
| register | **registrar** | right | **self** |
| short | **summer** | **time** | too |
| **weeks** | will | | |

**Table 1: The topics produced by *Balie*. The bold topics remain after additional filtering.**

## 3.3 Emotional Classification

Once the system has the list of entries and the associated topics, the next step is to generate data about the emotion expressed in each entry about each topic.

There are several types of existing systems that handle this task. One distinction among these systems is whether they determine only polarity [2, 26, 7, 8] or try to describe emotion more specifically [14, 23]. They also differ in whether they rely on some pre-determined description of individual words and sentences [2, 26], or whether they instead exploit the associations of larger texts and sentiment ratings [11, 10, 29].

Since our end goal is visualizing emotional states, there are multiple dimensions to the sentiment data we would like the system to generate. First, it should generate a valence score for each <entry, topic> pair. It should also identify the emotional context of each topic. To accomplish these tasks, we made use of an approach designed by Sood et al. called *Reasoning Through Search* [29]. The *RTS* classifier is trained on a corpus of product reviews that were accompanied by a numerical star rating. Given text, a subject on which to focus, and a method for determining context, the *RTS* system generates a query of nearby emotion-bearing words. Based on the number of appearances of the query words in positive or negative reviews, the system calculates a sentiment rating for the focus topic on the scale of -2 to 2 indicating the valence of the text with regard to the focus topic. This is the exactly the functionality needed to permit the generation of an emotional score given an <entry, topic> pair. The results of the sentiment query also provide the emotional context component we would like to visualize.

One useful feature of this system is the ability to specify the focus and focus method for the valence determination. The system provides several focus methods, including the sentence level as well as word windows of various sizes. The focus method determines which words are included in the sentiment query. The most accurate seems to be the "combination" focus method, which is described in detail in [30]. This method basically first attempts to apply the focus method with the highest accuracy, which happens to have limited coverage. If it lacks sufficient information to make the classification, the next most accurate method is applied. This process continues until a classification has been made or all the options have been exhausted.

In addition to using the final classification result, the system accesses the sentiment magnitude of the words that were used to make the sentiment classification [30]. The system then calculates the emotional rating of these words in isolation. This information provides the system with the option to display the most emotional words in the visualization.

## 3.4 EmoGraph

*EmoGraph* is the most basic of the visualization modes. It depicts the entire emotional trajectory of a topic's storyline using a standard graphing scheme.

### 3.4.1 User Scenario

Consider a user who feels like he is constantly struggling with emotional fluctuations about his financial situation. He wonders if his friends have had similar experiences. He could go through and reread each of his friends' previous blog entries and try to identify emotional changes related to their thoughts on money, but this would be tedious and time-consuming. VIBES provides an alternative. The user simply navigates to each of his friends' profile pages to see a list of topics important to each one. He chooses the topic, "money," and sees the following instance of *EmoGraph* (see Figure 4). He sees an emotional snapshot of each friend's "money" storyline. If he wants further detail on any one of the data points on the graph, he clicks on that point, which displays the entry corresponding to that point.

### 3.4.2 Design

For each topic, this visualization translates the valence and timestamp for each topical blog entry into a line graph (Figure 4). Time is represented along the x-axis and spans the entire range of relevant blog entries. In a future system, the user will be able to zoom in on a specific time range. The y-axis represents the emotion scale, from negative to positive. The scale of the y-axis ranges from -2 to +2, but since these numbers are not themselves revealing of the emotional status, we opted to use smiling and frowning icons to indicate the range of the y-axis. Previous work has found that such emoticons can be effective in communicating emotion [13]. Variation along the y-axis is depicted using a color gradient, where lightness corresponds to positive emotions and darkness to negative emotions. Each point on the graph corresponds to an entry. The end system will allow the user to click on these points to access the entries and overlay multiple topics on the same plot.

## 3.5 EmoMeters

### 3.5.1 User Scenario

Now consider a user who has recently discovered the blog of a long lost friend. She would like to quickly ascertain the current status of various aspects of her friend's life. She could try to read through all of the friend's blog entries from the past several months, but with VIBES, she does not need to. She can simply look at the *EmoMeters* display

...so right now i'm sitting in the computer lab with a late add form to force add recital attendance to this semester (i didn't register for it because i was under the impression i was finished with it) and i'm staking out the music office waiting for dr. gardner to return so i can have him sign this

form. then i need to turn it in to registrar and put all the remaining recitals in my planner so i can rearrange my schedule for them. i've refiled my intent to graduate and it now appears on my account. good. i don't know what happened to it the first time, i clearly remember doing it weeks ago, but whatever. it's there now. AAAAAAAAAAAAAAAUGH I JUST WANT TO GET OUT OF HERE! and then move and get married and teach. that's all i want right now. graduate, move, get married, teach. graduate, move, get married, teach. it's like a mantra that i will repeat to myself to keep me on track. *sigh* i feel much better now after ranting. i'm pretty sure i'll be able to handle all of this; i just needed to vent to get this all out. it'll be awkward, rearranging some scheduling and probably not being able to see chris as often (which REALLY sucks) but i'll do whatever it takes. long term goals, loyola. think long term goals. that's something you sort of suck at. *mutters to self* nothing can really be done, anyway, so just moral support and encouragement would be nice. okay i'm going to go. i need to keep an eye out for dr. gardner...
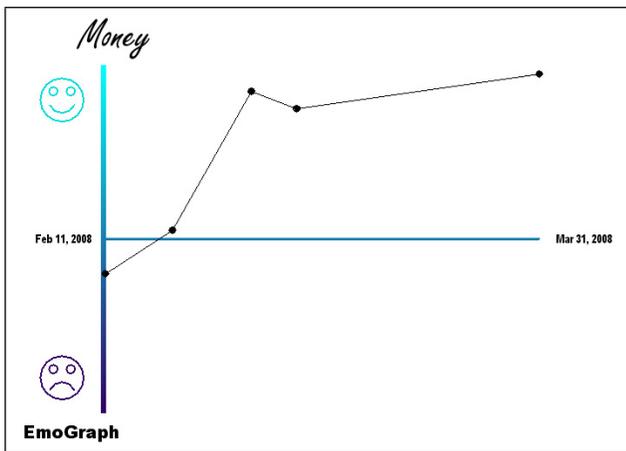
**Figure 3: Sample blog entry.**



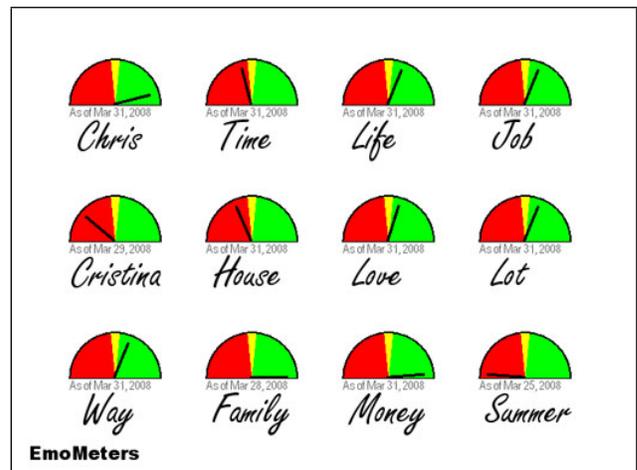**Figure 4: EmoGraph: changes in topical emotion over time.**



**Figure 5: EmoMeters: depicting current emotional state about a variety of topics.**

of her friend's profile (Figure 5), which depicts the most recent opinion expressed about each topic. If she would like to know more details, she can click on the topic to navigate to the relevant entry. She has now, at a glance, gleaned the most recent emotional content about the subjects most important to her old friend.

### 3.5.2  Design

The *EmoMeters* visualization sorts the topics according to number of entries in which they are mentioned. The rationale is that the more often a blog author discusses a topic, the more important it is to the author. In this visualization mode, the system then finds the most recent mention of each of the most important 12 topics. For each of these mentions, the valence is displayed in the form of a gauge. The angle of the gauge hand corresponds to the valence. There are three regions to each gauge: positive, negative, and neutral. These regions are colored according to popular mental associations between color and emotion. The positive region is green, the negative is red, and the neu-

tral is yellow. We decided on the gauge metaphor for this visualization since it conveys the current status of various aspects of the blogger's life. The twelve gauges are displayed together in the same space, which is reminiscent of a car's dashboard display. Unlike the other two visualizations, the *EmoMeters* visualization allows the user to make comparisons among topics. It also differs from *EmoGraph* in that it depicts static data rather than change over time.

## 3.6  EmoCloud

### 3.6.1  User Scenario

Suppose the previous user who has stumbled across the blog of an old friend and already perused the *EmoMeters* would like to know more about her friend's positive associations with "Chris." She simply needs to click on the corresponding *EmoMeter* to obtain an *EmoCloud* for the topic "Chris" (Figure 6). She will quickly be able to see both the negative and positive words her friend uses when discussing

Chris. This provides her with details about her friend's relationship with Chris without requiring her to read a large number of blog entries.



**Figure 6: EmoCloud: emotion bearing word associations**

### 3.6.2 Design

*EmoCloud* is inspired by the concept of tag clouds. The goal is to convey the nature of the emotional language used in reference each topic. EmoCloud is powered by the raw word-rating data from the emotional classifier. The system extracts the words used to make the sentiment query for each instance of the topic word. These are words located near the topic word that are used to determine the valence of the author's opinion about the topic. For each word in the query, the system retrieves the average of the probability that the word is negative and the probability that the word is positive based on the data from product review training in the *RTS* system. These words combine to generate a "sentiment magnitude" [29]. In preparation for the visualization, the system then filters words that do not meet a certain threshold, which was determined through experimentation. It is the words whose scores surpass this threshold that are the most emotional. The results of this process over all of a user's blog entries are combined. Duplicates are removed from the list, but the number of appearances of each word is stored. The system then classifies the valence of each of these words in isolation.

For each topic, the system generates a separate visualization. The topic word is displayed in the middle of the window. The emotion bearing words are displayed on either side. The positive words are displayed in blue under a plus sign on one side, and the negative words are in red under a minus sign on the other side. The size of each word is determined by the number of times it appears in the blog entries.

Like *EmoMeters*, *EmoCloud* displays a static state rather than change over time. A distinguishing characteristic of this visualization mode is its level of detail. While the other two visualizations depict only valence data, *EmoCloud* depicts a summary of overall valence as well as the actual words that influence this valence.

### 3.7 Summary

By identifying topics using Named Entity Recognition with additional filtering and classifying emotions using a system trained on product reviews [21, 29], VIBES generates topical emotion data that is displayed to the user through one of three visualizations. Users can share these visualizations through online profiles.

## 4. EVALUATION

Evaluating a system like VIBES is not a straightforward matter. We could test each component of the system by asking study participants specific questions about each topic presented and each emotion rating assigned. However, since the ultimate goal of the system is to generate a useful set of visualizations for the users and since testing each aspect of each visualization would be a considerable burden on the test users, it seemed the best way to evaluate the system was to use a holistic approach.

### 4.1 Methods

Of the 10 study participants, 40% were male and 60% were female. Half of the users were in the 25-30 age range, and the rest were distributed fairly evenly among 18-24, 30-35, and 40+. All of the participants had read and written in a blog at least once. Half of them said they read blogs several times a day. When asked about how much they write in blogs, 40% said they had only written in blogs a few times in their lives.

To test the system, we located a random blog using a built-in LiveJournal random blog selector [16]. VIBES processed this blog's entries from the past three months. We then asked the 10 participants to read through these blog entries, view the corresponding VIBES visualizations, and then rate their level of agreement or disagreement with a number of statements about the accuracy and utility of the visualizations generated by VIBES (see Table 2). The results of the study are displayed in Table 3[2].

### 4.2 Discussion

Participants generally agreed that the visualizations represented important topics, with $80-90\%$ agreeing somewhat or strongly with the first statement for each visualization. *EmoGraph* performed the best overall, with 70% of participants agreeing with statement 2 that the visualization betters understanding of emotional development. *EmoGraph* also got a 70% approval rating for being useful in the blogs the participants read.

*EmoMeters* and *EmoCloud* performed better at showing current state, with 60% in both cases agreeing with statement 3. Fewer participants, however expressed an interest in seeing *EmoCloud* or *EmoMeters* generated for the blogs they read or write.

The results regarding portrayal of current state versus change are expected given the differing goals of the various visualizations. The participant comments on the usefulness

---

[2]One participant opted out of responding to statement 5.

| Number | Statement |
|--------|-----------|
| 1 | This image represents a topic, Chris, that is important to this blogger. |
| 2 | This image helps me understand the blogger's emotional development. |
| 3 | This image accurately reflects the blogger's current attitudes toward this topic. |
| 4 | I would find it useful to see a similar image generated from topics in the blogs I read regularly (if you don't read blogs, suppose you did) |
| 5 | I would find it useful to see a similar image generated from topics in my blog (if you don't have a blog, suppose you did). |

**Table 2: Statements presented to study participants.**

| Statements | EmoGraph | EmoCloud | EmoMeters |
|-----------|----------|----------|-----------|
| 1 | 80% | 90% | 80% |
| 2 | 70% | 40% | 30% |
| 3 | 50% | 60% | 60% |
| 4 | 70% | 40% | 40% |
| 5 | 50% | 20% | 33.3% |

**Table 3: Percentage of participants that agreed somewhat or strongly with each statement.**

of the visualizations will be helpful in refining the system. For example, it seems a more explanatory interface is required for *EmoMeters*.

## 5. FUTURE WORK

Potential extensions to VIBES include the addition of interactivity, including topic and date selection as well as navigation capabilities. In addition, future innovations in topic identification and emotion classification can easily be integrated into VIBES because of its modularity.

There are also a number of avenues for exploration into the application of VIBES technology in emotional intelligence research. We are currently investigating the similarity in emotional pattern data among different bloggers. For example, once integrated into a social blogging site, VIBES could be extended to allow bloggers to search for people whose emotional patterns are similar to their own. We define similarity in terms of the closeness of vectors that summarize each emotion data point.

Another extension involves comparing these vectors among different topics to analyze which topics are associated with related emotional patterns.

## 6. CONCLUSIONS

We have developed a system called VIBES, which is intended to further our understanding of emotional intelligence by visually depicting the evolving emotional states in blogs at the individual blogger level. We have combined topic identification, emotion classification, and visualization technologies in order to accomplish this task.

VIBES generates three visualizations: *Emograph*, which depicts changes in emotional valence throughout topic-specific storylines; *EmoCloud*, which displays topical words with the

greatest emotional intensity; and *EmoMeters*, which provides a status display of emotions with regard to topics most important to a blogger.

As a majority (70%) of study participants found *EmoGraph* helpful in understanding the emotional development of the blogger, we are confident that the system has potential to expand the emotional intelligence of blog readers. The user tests also suggest that changes in emotion over time are more useful to users than depictions of static state. Additional user testing is required to determine the usefulness to blog writers. The *EmoCloud* and *EmoMeters* displays were considerably less popular among participants, suggesting that improvements in clarity and aesthetic qualities may be required before the system could be deployed.

Finally, the *EmoGraph* visualization of VIBES provides a framework for modeling emotional variation over time. This framework can help researchers develop emotionally intelligent interfaces that can sense and respond to human emotions, which remains an open problem in affective computing.

## 7. REFERENCES

[1] C. H. Brooks and N. Montanez. An analysis of the effectiveness of tagging in blogs. In *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*, 2006.

[2] X. Ding and B. Liu. The utility of linguistic rules in opinion mining. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 811–812, New York, NY, USA, 2007. ACM.

[3] Golan Levin and Kamal Nigam and Jonathan Feinberg. The dumpster. `http://artport.whitney.org/commissions/thedumpster/`, 2005.

[4] D. Goleman. *emotional intelligence*. Bantam, 1995.

[5] Z. Jingbo and Y. Tianshun. A knowledge-based approach to text classification. In *Proceeding of the first SIGHAN workshop on Chinese language processing*, pages 1–5, Morristown, NJ, USA, 2002. Association for Computational Linguistics.

[6] Jonathan Harris and Sep Kamvar. We feel fine. `http://www.wefeelfine.org/`, 2005.

[7] J. Kamps and M. Marx. Words with attitude. In *1st International WordNet Conference*, pages 332–341, 2002.

[8] S.-M. Kim and E. Hovy. Determining the sentiment of opinions. In *COLING '04: Proceedings of the 20th international conference on Computational Linguistics*, page 1367, Morristown, NJ, USA, 2004. Association for Computational Linguistics.

[9] L.-W. Ku, L.-Y. Lee, T.-H. Wu, and H.-H. Chen. Major topic detection and its application to opinion summarization. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 627–628, New York, NY, USA, 2005. ACM.

[10] D. Kushal, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 519–528, New York, NY,

USA, 2003. ACM.

[11] G. Leshed and J. J. Kaye. Understanding how bloggers feel: recognizing affect in blog posts. In *CHI '06: CHI '06 extended abstracts on Human factors in computing systems*, pages 1019–1024, New York, NY, USA, 2006. ACM.

[12] H. Li and K. Yamanishi. Topic analysis using a finite mixture model. In *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora*, pages 35–44, Morristown, NJ, USA, 2000. Association for Computational Linguistics.

[13] H. Liu, H. Lieberman, and T. Selker. A model of textual affect sensing using real-world knowledge. In *IUI '03: Proceedings of the 8th international conference on Intelligent user interfaces*, pages 125–132, New York, NY, USA, 2003. ACM.

[14] H. Liu, T. Selker, and H. Lieberman. Visualizing the affective structure of a text document. In *CHI '03: CHI '03 extended abstracts on Human factors in computing systems*, pages 740–741, New York, NY, USA, 2003. ACM.

[15] Livejournal. About livejournal. `http://www.livejournal.com/site/about.bml`, 2008.

[16] LiveJournal. Random journal. `http://www.livejournal.com/random.bml`, 2008.

[17] Livejournal. Statistics. `http://www.livejournal.com/stats.bml`, 2008.

[18] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai. Topic sentiment mixture: modeling facets and opinions in weblogs. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 171–180, New York, NY, USA, 2007. ACM Press.

[19] G. Mishne and M. de Rijke. Deriving wishlists from blogs show us your blog, and we'll tell you what books to buy. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 925–926, New York, NY, USA, 2006. ACM.

[20] G. Mishne and M. de Rijke. Moodviews: Tools for blog mood analysis. In *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*, 2006.

[21] D. Nadeau. *Semi-Supervised Named Entity Recognition*. PhD thesis, Ottawa-Carleton Institute for Computer Science, Ottawa, CA, 2007.

[22] Nielsen BuzzMetric. Blogpulse. `http://www.blogpulse.com`, 2008.

[23] J. Oberlander and S. Nowson. Whose thumb is it anyway?: classifying author personality from weblog text. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 627–634, Morristown, NJ, USA, 2006. Association for Computational Linguistics.

[24] Pew Internet and American Life Project. Teens and social media, 2007.

[25] Pew Internet and American Life Project. Internet activities, 2008.

[26] A.-M. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 339Ű–346. Association for Computational Linguistics, 2005.

[27] J. Schmidt. Blogging practices: An analytical framework. *Journal of Computer-Mediated Communication*, 12(4), 2007.

[28] S. Smale and S. Greenberg. Transient life: collecting and sharing personal information. In *OZCHI '06: Proceedings of the 20th conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: activities, artefacts and environments*, pages 31–38, New York, NY, USA, 2006. ACM Press.

[29] S. Sood, S. Owsley, K. Hammond, and L. Birnbaum. Reasoning Through Search: A Novel Approach to Sentiment Classification. *submitted to EMNLP, July*, 2006.

[30] S. O. Sood. *Compelling Computation: Strategies for Mining the Interesting*. PhD thesis, Northwestern University, Evanston, IL, USA, 2007. Adviser-Kristian Hammond.

[31] M. Willis. An emotionally intelligent user interface: modelling emotion for user engagement. In *OZCHI '07: Proceedings of the 2007 conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: activities, artifacts and environments*, pages 187–190, New York, NY, USA, 2007. ACM.

[32] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM-2003)*, pages 427–434, 2003.