Pomona College
Department of Computer Science

# VIBES: Visualizing Changing Emotional States in Blogs

April Wensel

May 1, 2008

Submitted as part of the senior exercise for the degree of
Bachelor of Arts in Computer Science

Professor Sara Owsley Sood, adviser

## Abstract

By creating new modes of interaction between humans, the Internet provides new opportunities for the exchange of emotional experiences. Online journals (blogs), for example, provide not only an outlet for emotional self-expression, but also a space for social interaction and commiseration. However, the massive extent of the blogosphere can overwhelm users with data and restrict their ability to make meaningful connections to fellow bloggers. Recent research into this problem has focused on methods of extracting global trends in opinion across the entire blogosphere, but there has been considerably less work on automatically summarizing content from bloggers individually. Therefore, a need exists for a system that can organize and summarize personal blog data into a manageable, accessible format.

We have developed such a system, VIBES, that extracts the most important topics from a blog, measures the emotions associated with those topics, and generates a suite of visualizations that summarize those emotions and how they evolve over time. In preliminary user tests, a majority of participants said they would find it useful to see at least one of the visualizations for the blogs they read or write. Most also agreed that each visualization either revealed the author's current emotional state or emotional development over time. VIBES has potential applications both in sharing emotional profiles among online friends 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.

# Contents

# List of Figures

# List of Tables

# Chapter 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, which, respectively, relate to the understanding of one's own emotions and the emotions of others [Gol95]. Many believe that these skills are essential to success in life.

> *If your emotional abilities aren't in hand, if you don't have self-awareness, if you are not able to manage your distressing emotions, if you can't have empathy and have effective relationships, then no matter how smart you are, you are not going to get very far.*

Daniel Goleman [Den04]

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" [Sch07]. For the blogger, therefore, the expression of emotion can be cathartic and may evoke empathy in the form of reader responses.

*To give vent now and then to his feelings, whether of pleasure or discontent, is a great ease to a man's heart.*

Francesco Guicciardini [Gui90]

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 [Pew08]. The number of blog authors is greater among online teenagers, with 28% keeping a blog [Pew07].

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 size of the ever-expanding blogosphere makes it impossible for readers to locate all of the entries that may have emotional relevance to them. The problem is similar for writers. While they may benefit from expressing their emotions, it is difficult for them to step back and construct 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 research 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 exchange of emotional experiences among people. By extracting emotional content automatically and presenting it to the user in a useful and meaningful way, VIBES can help users overcome information overload and access specific emotional content from blogs. Rather than seeking to represent the *vox populi* as other systems have, VIBES will preserve and amplify the voice of the individual.

*Your feelings express the deepest truth of your soul...Feelings are the universal human language, a conduit from heart to heart, transcending our outer differences and connecting us to all.*

Judith Wright [Wri06]

VIBES generates visualizations that depict the emotional content from an individual blog in three different ways. These visualizations can be placed

[1] Visualizations of Individual Bloggers' Emotional States

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. The practice of displaying the results of such quizzes publicly—in addition to the very practice of blogging itself—indicates bloggers' willingness to share this information. This phenomenon may be due a combination of the author's narcissism and the reader's voyeurism, but a less cynical perspective might explain the practice as part of the neverending quest for people to understand themselves and how they relate to the rest of the world.

In addition, the visualizations offer a tool with which researchers can model "emotional state transition" [Wil07]. Knowledge of these emotional patterns can assist in building emotionally intelligent interfaces that sense and respond to emotional fluctuations in the user.

# Chapter 2

# Background

Three underlying technologies are required to build a system like VIBES. There must be a way to determine the important topics from a blog, measure the author's emotion toward those topics, and display the results effectively to the user. There have been relevant previous studies in each of these areas, several of which have directly or indirectly influenced the construction of VIBES.

## 2.1 Topic Identification

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 while the second does not.

### 2.1.1 Topic Classification

Topic classification systems determine relevant topics probabilistically from a given set of potential topic words. Implementations of these systems often involve using machine learning to train the systems to recognize text with a certain topic based on a corpus of texts tagged with topics. These learning techniques, which include Hidden Markov Models, Support Vector Machines, and Naïve Bayes Classifiers, generally work by identifying related words in the corpus and then searching for these in the unclassified texts [JT02, LY00, MLW$^+$07].

VIBES poses a unique challenge because these 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. One book recommendation system uses the list of interests listed in bloggers' profiles as the list of potential topics [TC06]. We decided against using interests listed in profiles as the potential topics for several reasons. Many bloggers list few or no interests, and those that do list interests tend to focus on common topics such as musicians or favorite foods. One of the main goals of listing interests, after all, is to provide a common ground between bloggers. Therefore, interests supplied by the blogger do not necessarily reflect the topics an individual writes about frequently. For example, while a person's name may represent a very imporant topic to the blogger, it is unlikely to be listed in the blogger's interests. In the absence of a list of candidate topics, traditional training methods would not apply. Therefore, we did not model VIBES on systems that use techniques relying on knowing a topic or group of topics ahead of time.

### 2.1.2 Keyword Extraction

Keyword-based approaches, on the other hand, can overcome this lack of previous knowledge. These approaches rely on characteristics of the text itself to extract important words. One such characteristic is the relative frequencies of words in a given document compared to others. Tagging suggestion and product recommendation systems for blogs often use this approach [BM06, MdR06a].

Other previous systems build upon this basic technique by considering semantic relationships among words. Ku et al., for example, built a system that finds likely candidates for keywords, enumerates related terms, and then searches for sentences in the text that contain a significant number of the relevant words to determine the topics in the document [KLWC05]. Other systems have applied similar word distribution models to identify topics [MLSZ06, BEYTW03, WKY96].

A final variation on the keyword method is Named Entity Recogntion, which extracts entities from text and classifies them into categories such as people, places, and organizations. One such system is *Balie*, which uses a semi-supervised learning technique to train a system to identify named entities [Nad07]. *Balie* is especially impressive because it requires only a short list of training examples, and relies on search to learn to extract entities. Since *Balie* does not require prior knowledge of topics, and since it identifies proper nouns, which are likely to represent topics important to a blogger, we decided to use this system as part of our topic identification scheme. The details are discussed further in Chapter 3.

## 2.2 Emotion Classification

Following the topic identification step, the system must classify the blogger's emotions about those topics. There are several types of existing systems that handle this task. One distinction among these systems is whether they determine only polarity or try to describe emotion more specifically. They also differ in whether they rely on some pre-determined description of individual words and sentences, or whether they instead exploit the associations of larger texts and sentiment ratings.

One category of approaches relies on the linguistic properties of the text in order to determine the valence of the language used. Kim and Hov created a system that determines sentiment at both the word and sentence level in order to locate individuals that hold a specified opinion [KH04]. The system's classifier relied on generating lists of positive and negative words using *WordNet* and a small number of seed words. Kamps and Marx have implemented a similar approach that uses "good" and "bad" as seed words [KM02].

Ding and Liu argue that pure linguistic analysis is not enough, and that the semantic orientation of words in context must also be considered [DL07]. Popescu and Etzioni describe a system that similarly uses the semantic orientation of words in order to mine product reviews from documents [PE05].

Other systems seek to represent emotions in more detail than a simple determination of positive or negative sentiment. Liu et al. take a more complex, multilayered approach, building a system that first senses emotion at the sentence level and then groups sentences together, essentially generating an "affective structure" of the document [LSL03]. Their system relies on commonsense knowledge from the *Open Mind Common Sense* to classify sentences according to the six classic emotions enumerated by Ekman: happy, sad, angry, fearful, disgusted, and surprised, or neutral. Their work is unique in that it can sense emotions that word-based classification would not. The example they provide is that their system can tell that the sentence, "My wife left me; she took the kids and the dog," is sad while a simple word classification scheme likely would not.

Some research has attempted the in-depth classification of the personality and opinions of bloggers than a simple negative or positive distinction [ON06]. They attempt to classify blog text according to such personality dimensions as agreeableness, conscientiousness, extraversion, and neuroticism. Their focus on individual blogger personalities is similar to VIBE's goal of preserving the identity of the individual; however, they are more focused on static qualities of the blogger's personality than change over time.

Another group of sentiment classification techniques takes a more holistic approach, basing decisions on training data that associates a mood or emotion with a whole text. Leshed and Kaye, for example, used a machine-learning technique to automatically categorize text by the emotion of the author [LK06]. They used a corpus of blog posts tagged for mood to train a support vector machine according to the word frequency and ignoring word order. In discerning specific moods, their approach resulted in high accuracy and precision, but low recall. Their approach differs from others in that they did not focus on the pre-defined meanings of words, and also, they were more concerned with predicting the authors' perceptions of their emotions rather than some absolute emotional quality of the text.

There are a number of similar systems that use customer reviews in order to train a system. These systems look for associations between the appearance of certain terms in a review and the numerical score the rater assigned. One such system applies an approach called *Reasoning Through Search*, which determines the sentiment score of words based on their appearance in positive or negative reviews [SOHB06]. Some of these systems look for text classified as either positive or negative, and others classify the strength of the sentiment. Kushal et al., for example, discuss a feature extraction and scoring system to classify customer reviews found in arbitrary documents online [KLP03]. They trained their system on reviews that were tagged by the reviewer with a positive or negative rating.

It seems that the most successful approaches to emotion classification incorporate both linguistic and contextual data. For VIBES, we chose to apply the *Reasoning Through Search* approach since it will provide the system with data on the valence and intensity of emotion.

## 2.3 Visualization

The final component of the system involves generating an intuitive and informative visualization of the topic and emotion data assembled by the system. Previous work has examined ways of representing the topics in a document, and other work has explored ways of depicting emotions. There are a few studies that attempt to represent both topic and opinion data, and others still that can depict changes over time. There are also several goals for such visualization systems.

One set of goals relates to representing document structure. Liu, Selker, and Lieberman designed a system that visualizes the emotional sections of a story document with an interactive color bar. The user can navigate to parts of the document with a specific emotion using this bar. In user tests, they found that their system improved navigation speed [LSL03]. Qi and Candan briefly discuss a method for visualizing topics using gradient-filled rectangles whose color and size reveals information about the document segment's length, topic concentration, and topic variation [QC06].

Many visualization projects attempt to communicate opinions on products. Liu, Hu, and Cheng have designed a visualization that displays the positive and negative scores for features of a set of products [LHC05]. The visualization consists of several rectangles displayed on a graph where the vertical axis ranges from negative to positive. The horizonal axis lists the various features of a given product. For each feature, there are several rectangles, each of which corresponds to a brand of the product. The area of the rectangle above the horizontal axis corresponds to the percentage of positive reviews for that the specified feature of that brand. Similarly the area below the axis represents the percentage of negative reviews. Their representation is static, while VIBES intends to represent change over time.

The system described by Gregory et al. produces several visualizations of topic-specific opinion data about products [GCW+06]. One of them is a rose plot whose petals correspond to depictions of positive vs. negative, pleasure vs. pain, cooperation vs. conflict, and vice vs. virtue. They also describe a galaxy visualization that displays user reviews clustered by thematic content. They also show a simpler tool that displays the affective rating of a given set of topics, which also has the option of showing change over time.

Another goal of visualization is simply to reflect the emotions expressed in interactions among users. One such visualization is part of a tool called EmpathyBuddy, which determines the sentiment of an email message and generates an animated face that depicts the corresponding emotion [LLS03]. The aim is to communicate the emotion conveyed by the text to the writer and the reader.

Appan, Sundaram and Tseng have designed a novel ring visualization that depicts information about topic discussion in social networks over time [AST06]. Since their work focused strictly on the visualization, they decided topics based solely on mention of the topic in message subjects. Their tool consists of a number of concentric circles. The timescale is from the innermost ring to the outermost ring. Colored dots correspond to people who have discussed a given topic. In user tests, the investigators found that users were interested in "visualizing relationships between a single person time and the topic" as well "as comparing communication activity for multiple topics in the same visualization" [AST06]. VIBES attempts to accomplish both of these tasks in addition to adding the emotional component.

VIBES is particularly inspired by Mishne and de Rijke's MoodViews, which relies on simple graph plots to depict change in global mood level in blog posts over time. The date is on the horizontal axis and the number of people with a given mood are on the vertical axis [MdR06b]. Moodviews and other related end systems are discussed further in Chapter 5.

The visualization goals of VIBES include summarizing opinions with regard to specific topics, illustrating change over time, and representing the individual. Therefore, we have incorporated aspects from a variety of these related visualization schemes.

## 2.4 Summary

There is clearly a substantial history of work done in each of the components of the system. To expedite the creation of the system, and so that the focus could be on combining and representing the resulting information in an interesting and useful way, existing technologies are used in VIBES where appropriate and adapted to suit the specific needs of the project. Specifically, the variability of blog data led us to adapt *Balie* for topic identification and *Reasoning Through Search* for emotion classification.

# Chapter 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 3.1). 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 3.1. 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 and generates three visualizations.

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 is that it is formatted for 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

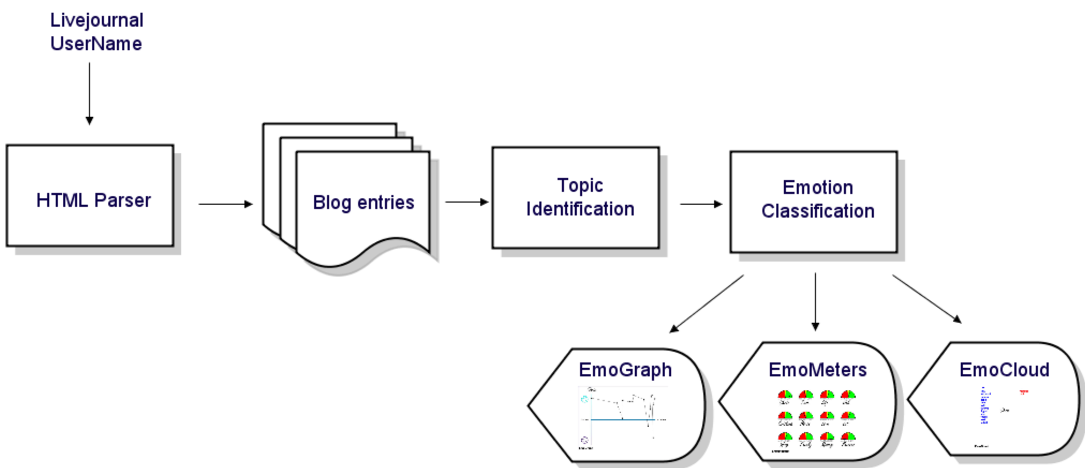


Figure 3.1: High-level system architecture.

as "<pubDate>" and "<title>." However, these feeds tend to be limited to only the most recent 25 or so 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 [Liv08a] for a number of reasons. First of all, LiveJournal is one of the most popular blogging sites, with 15,459,629 users total [Liv08c]. 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 [Liv08a]. Finally, LiveJournal has a fairly consistent format for displaying entries. Most LiveJournal navigation and 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 entries.

The entry extractor automatically retrieve entries for a specified blogger by constructing appropriate URLs and extracting only the blog text from the HTML. Specifically, the 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 owned the blog. An example URL is "http://username.livejournal.com/2000/03/11/." 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. On each entry page, the parser first searches for the subject of the blog post and then retrieves the text following the title, storing it as the entry text.

## 3.2 Topic Identification

The next step entails identifying the topics in each of the entries. The overall structure of this module is depicted in Figure 3.2.

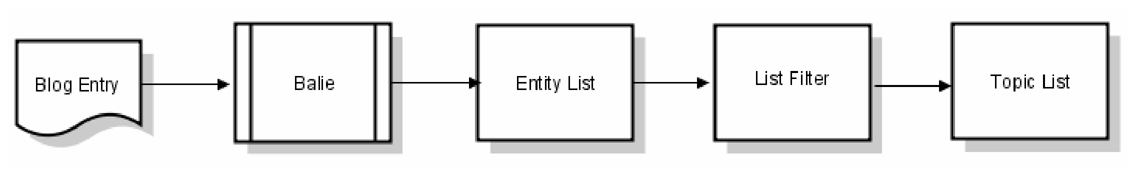


Figure 3.2: Topic identifier module.

At the heart of the topic identifier is a system called *Balie*, which identifies named entities in a novel semi-supervised way [Nad07]. *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 unannotated 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.

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 entity through a filter (Figure 3.3). The filter first checks to make sure

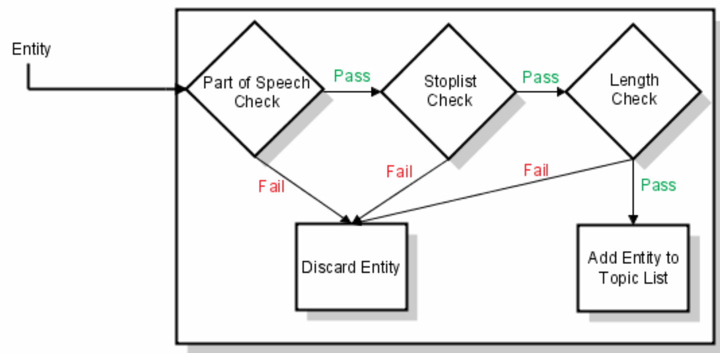


Figure 3.3: Individual entity filter.

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 [PE05, YNBN03]. 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 3.2 displays a blog entry excerpt, the rest of which can be found in Appendix A. Table 3.1 illustrates the results of the topic filtering process.

## 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. Since our end goal is visualizing emotional states, there are multiple dimensions to the sentiment data we would like the system to gen-

...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 3.4: Sample blog entry.

| 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 3.1: The topics produced by *Balie*. The bolded blue topics remain after additional filtering.

erate. 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* [SOHB06]. 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 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. There are several focus methods possible, 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 choice of focus seems to be the "combination" focus method, which is described in detail in [Soo07]. 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 [Soo07]. 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 emotional change about a topic over time 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 profile page of his friends to see a list of topics important to each one 3.5. He chooses the topic, "money," and sees the following instance of EmoGraph. He looks at the ups and downs of each friend's emotions relating to money. If he wants further detail on any one of the points, he clicks on the dot at that point, which displays the entry corresponding to that point.

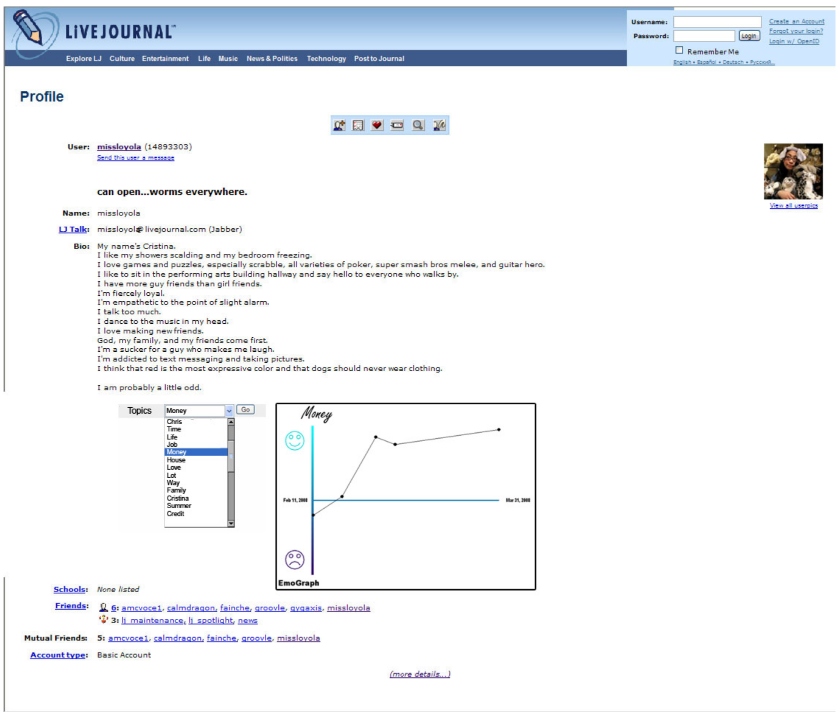


Figure 3.5: Mockup of VIBES integrated into a LiveJournal user profile. [Cri08]

### 3.4.2 Design

For each topic, this visualization translates the valence and timestamp for each topical blog entry into a line graph 3.6. Time is represented along the x-axis and spans the entire range of relevant blog entries. In the future, 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 to some extent arbitrary and 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 [LLS03]. 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.

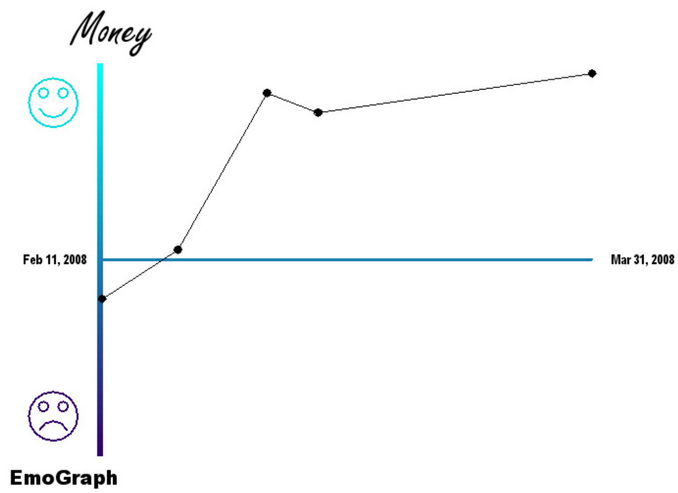


Figure 3.6: EmoGraph: changes in topical emotion over time.

## 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 of her friend's profile (Figure 3.7), 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.

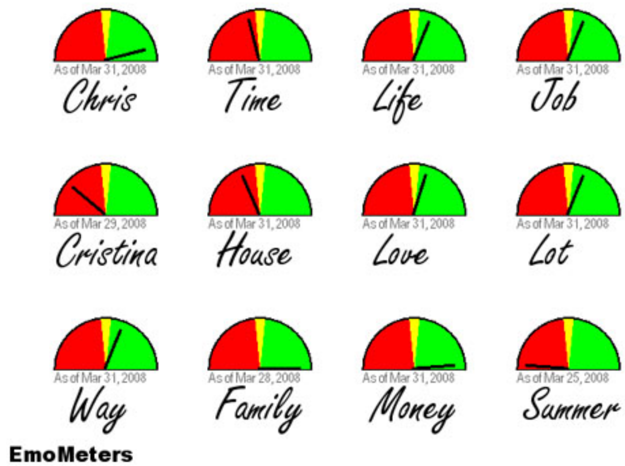


Figure 3.7: EmoMeters: depicting current emotional state about a variety of topics.

### 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. 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 neutral 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, EmoMeters 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 a few more details about her friend's highly positive associations with Chris. She simply needs to click on the corresponding EmoMeter in order to obtain an EmoCloud for the topic "Chris." 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.

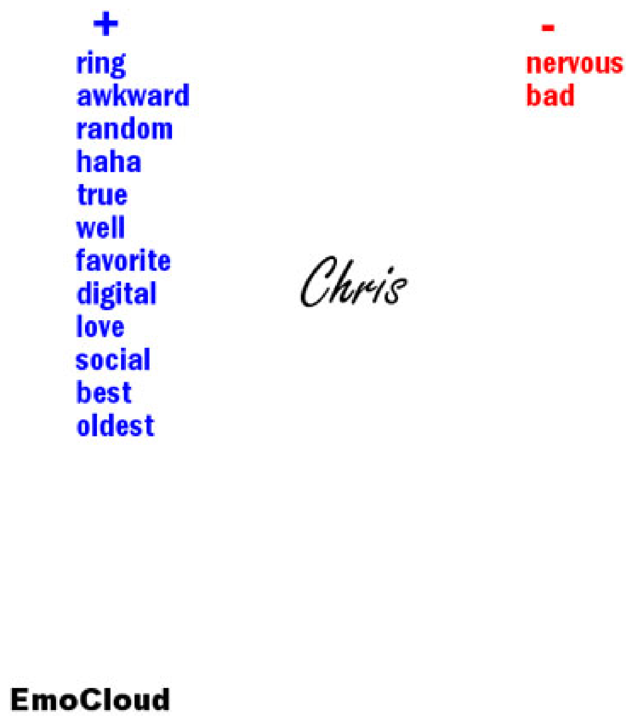


Figure 3.8: EmoCloud: emotion bearing word associations

### 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 about 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 make the valence calculation. 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" [SOHB06]. 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, EmoClouds 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, EmoClouds 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 based on product reviews [Nad07, SOHB06], VIBES generates topical emotion data that is displayed to the user through one of three visualizations. Users can share these visualizations through online profiles.

# Chapter 4

# Evaluation

Evaluating a system like VIBES is not a straightforward matter. We could test each component of the system by asking 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 Demographics

Of the 10 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, and 40% said they had only written in blogs a few times in their lives.

## 4.2 Methods

To test the system, we located a random blog using a built-in LiveJournal random blog selector [Liv08b]. VIBES processed three months of entries from this blog. We then asked the 10 participants to read through these blog entries 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. The results of the study are displayed in Table 4.1.

| | EmoGraph | EmoCloud | EmoMeters |
|---|---|---|---|
| 1. This image represents a topic, Chris, that is important to this blogger. | 80% | 90% | 80% |
| 2. This image helps me understand the blogger's emotional development. | 70% | 40% | 30% |
| 3. This image accurately reflects the blogger's current attitudes toward this topic. | 50% | 60% | 60% |
| 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). | 70% | 40% | 40% |
| 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). | 50% | 20% | 33.3% |

Table 4.1: Percentage of participants that agreed somewhat or strongly with each statement. (The 33.3% in the EmoMeters column is a result of one participant opting out of responding to statement 5.)

## 4.3 Anecdotal Evidence

We also collected open-ended feedback on each of the visualizations, which will be helpful as refine the system.

### 4.3.1 EmoGraph

Participants found the EmoGraph provided an alternate perspective for understanding blog entries. They provided the following comments.

> *It's interesting to see how her feelings/opinions shift throughout the three months. The graph gives you a direct view of how the blogger felt throughout the months, whereas by reading the blog entries, you may not have noticed how the blogger's feelings had changed.*

> *They really give you a feel of how the blogger is feeling, pictures say alot and I myself personally love when a blog displays pictures.*

> *The image helps follow the writer's emotional states without reading the entries.*

### 4.3.2 EmoCloud

Participants expressed an interest in seeing the details provided by EmoCloud, and they as noted the lack of temporal data.

> *This seems more useful for knowing about how someone feels about something overall; I don't get a sense of temporal change, however.*

> *The image helps show how positive the writer is feeling, but does not give any indication of time.*

> *Images say alot and they, to me give me an idea about the reader. I do like the headline's they really captivate me to want to read the story.*

### 4.3.3 EmoMeters

Participant comments on EmoMeters suggest that a reworking of the visualization could convey the information more intuitively.

> *The image helps show the writer's emotional state pertaining to various subjects.*
>
> *The text in the image was not very clear for me personally, I personally dont care for this.*
>
> *The image helps show the writer's emotional state pertaining to various subjects.*
>
> *Seems vague*
>
> *I am a little confused as to how the meters work*

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

# Chapter 5

# Related Systems

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 [Nie08]. 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 [MdR06b].

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 [Jon05]. A similar project is *The Dumpster*, which does a similar text search for descriptions of romantic turmoil [Gol05]. The surrounding text is analyzed in order to group breakups 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 [SG06]. 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.

# Chapter 6

# Future Work

There are several refinements required before VIBES would be ready for deployment. There are also a number of avenues for exploration into the application of VIBES technology in emotional intelligence research.

## 6.1 Interaction and Deployment

Some of the interesting extensions revolve around making the system interactive. While our user tests focused on the evaluation of static images, the ultimate goal of the system is to provide an interactive format for managing the emotional content of blogs. Therefore, one type of interaction will be the ability to navigate blog entries through the visualizations. For EmoGraph, this would first mean an additional control panel listing all of the topics and linking them to the individual instances of EmoGraph. In addition, each of the blog entry points on the graph would be linked to the corresponding entry. It would also be useful for mouse hovering to display the date and title of the entry. The EmoMeters visualization would contain links to the most recent entry on each topic. EmoCloud would have the same topic control panel as EmoGraph, and it would also provide links from each emotion bearing word to the entry in which it appears.

Each visualization should also provide a method for changing the date window depicted by the image. Also, users should be able to overlay their visualizations with corresponding instances of other users' visualizations.

The final step would be actually integrating VIBES into bloggers' user profiles. This would require setting up a web interface for the system, through which an arbitrary blogger URL can be entered. The final system should also provide support for other blogging sites than LiveJournal.

This will simply require a more sophisticated parser.

## 6.2 Modeling Emotional Patterns

A separate area for exploration is measuring the similarity in data among different bloggers. One approach would be to identify distinguishing features of various EmoGraph instances. These could include information about the shape of the function. For example, this could include whether it is increasing or decreasing, how many peaks it has, and how extreme the differences are between negative and positive points. It could also include a measure of variability. Another interesting feature would be whether the topical entries overall are mostly positive or negative.

Once a set of features is determined, one could weight each feature in the determination of a similarity score between instances of the EmoGraph data. This similarity rating would facilitate the creation of a number of extensions to the VIBES interface. Specifically, users would be able to, with the click of a button, find users who have exhibited a similar emotional pattern, either in general or regarding a specific topic. This data could also enable the creation of emotion "forecasts" that make prediction about emotional trajectories based on how similar patterns evolved.

In addition to use in VIBES, a suitable quantification of similarity in emotional patterns would also facilitate the creation of a sentiment-based search engine. This could help users find people with similar emotions as quickly as they find webpages with Google. If blog search really is only at the level of webpage search in 1997 [NXL+07], then new navigation strategies involving sentiment are worth exploring.

Quantification of the relationship has further applications in the identification of global classes of emotional patterns. For example, a system may be able to identify a specific emotional trajectory such as the stages of grief or the emotions following the end of a relationship.

Finally, identifying emotional patterns would also be useful for research in creating adaptive affective interfaces. To produce machines that can exhibit human emotion, it is necessary to understand how these emotions change over time. Consider a chat bot that provides technical assistance. It would benefit from being able to detect an emerging emotional pattern in the user so that it can in turn, adapt its emotional response.

## Chapter 7

# Conclusions

Emotional intelligence is essential to being human as well as simulating humans. For this reason, we have developed a system called VIBES, which is intended to further the cause 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 with regard to specific topics over time; 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.

The fact that a majority (70%) of study participants found EmoGraph helpful in understanding the emotional development of the author provides promising evidence 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.

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.

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.

# Appendix A

# Sample Blog Entry

ok. so today i went in for a degree audit just to make sure that i was on the right track to graduation. good thing i did, too, because there were two errors: 1. my intent to graduate apparently was not registering under my account, so the school had no idea i was graduating this summer 2. three semesters of my recital attendance were one recital short - and i didn't get credit for them, even though i have been under the impression that i had. let me explain something about recital attendance. it is a 0 credit course that every music student has to take, and you have to attend and turn in proof of attendance for 10 recitals throughout the course of the semester. it sounds pretty easy, however, most recitals take place during the day (while classes are in session) and when i've worked, most have also been during working hours. it really irritates me that i can't just make up the remaining THREE recitals to complete those semesters. instead, in order to graduate, i have to attend THIRTY recitals by the end of May. THIRTY. there are apparently over fifty scheduled, but i work two jobs, sub at another, and i'm taking three classes. so i guess you could say i'm a little stressed out. it further irritates me that i didn't know about any of this until today. i had checked my online account, i had talked to my advisors, everything. and i didn't know about this until today?! i have been talking to people about this for at least three semesters. if i had known about all this i could have taken recital attendance those three semesters and made life a hell of a lot easier for me now. 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. <3 [Cri08]

# Bibliography


[AST06] Preetha Appan, Hari Sundaram, and Belle L. Tseng. Summarization and visualization of communication patterns in a large-scale social network. In *PAKDD*, pages 371–379, 2006.

[BEYTW03] Ron Bekkerman, Ran El-Yaniv, Naftali Tishby, and Yoad Winter. Distributional word clusters vs. words for text categorization. *J. Mach. Learn. Res.*, 3:1183–1208, 2003.

[BM06] Christopher H. Brooks and Nancy Montanez. An analysis of the effectiveness of tagging in blogs. In *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*, 2006.

[Cri08] Cristina Loyola. can open...worms everywhere. `http://missloyola.livejournal.com`, 2008.

[Den04] Dennis Hughes. Interview: Daniel goleman on emotions and your health. `http://www.shareguide.com/Goleman.html`, 2004.

[DL07] Xiaowen Ding and Bing 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.

[GCW+06] Michelle L. Gregory, Nancy Chinchor, Paul Whitney, Richard Carter, Elizabeth Hetzler, and Alan Turner. User-directed sentiment analysis: Visualizing the affective content of documents. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, pages 23–30. Association for Computational Linguistics, 2006.

[Gol95] Daniel Goleman. *emotional intelligence*. Bantam, 1995.

[Gol05] Golan Levin and Kamal Nigam and Jonathan Feinberg. The dumpster. `http://artport.whitney.org/commissions/thedumpster/`, 2005.

[Gui90] Francesco Guicciardini. *Counsels and Reflections of Francesco Guicciardini*. Kegan Paul, Trench, Trübner and Co., 1890.

[Jon05] Jonathan Harris and Sep Kamvar. We feel fine. `http://www.wefeelfine.org/`, 2005.

[JT02] Zhu Jingbo and Yao 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.

[KH04] Soo-Min Kim and Eduard 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.

[KLP03] Dave Kushal, Steve Lawrence, and David 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.

[KLWC05] Lun-Wei Ku, Li-Ying Lee, Tung-Ho Wu, and Hsin-Hsi 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.

[KM02] Jaap Kamps and Maarten Marx. Words with attitude. In *1st International WordNet Conference*, pages 332–341, 2002.

[LHC05] Bing Liu, Minqing Hu, and Junsheng Cheng. Opinion observer: analyzing and comparing opinions on the web. In *WWW '05: Proceedings of the 14th international conference on World Wide Web*, pages 342–351, New York, NY, USA, 2005. ACM.

[Liv08a] Livejournal. About livejournal. `http://www.livejournal.com/stats.bml`, 2008.

[Liv08b] LiveJournal. Random journal. `http://www.livejournal.com/random.bml`, 2008.

[Liv08c] Livejournal. Statistics. `http://www.livejournal.com/stats.bml`, 2008.

[LK06] Gilly Leshed and Joseph 'Jofish' 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.

[LLS03] Hugo Liu, Henry Lieberman, and Ted 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.

[LSL03] Hugo Liu, Ted Selker, and Henry 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.

[LY00] Hang Li and Kenji 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.

[MdR06a] Gilad Mishne and Maarten 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.

[MdR06b] Gilad Mishne and Maarten de Rijke. Moodviews: Tools for blog mood analysis. In *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*, 2006.

[MLSZ06] Qiaozhu Mei, Chao Liu, Hang Su, and ChengXiang Zhai. A probabilistic approach to spatiotemporal theme pattern mining on weblogs. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 533–542, New York, NY, USA, 2006. ACM.

[MLW+07] Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang 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.

[Nad07] David Nadeau. *Semi-Supervised Named Entity Recognition*. PhD thesis, Ottawa-Carleton Institute for Computer Science, Ottawa, CA, 2007.

[Nie08] Nielsen BuzzMetric. Blogpulse. `http://www.blogpulse.com`, 2008.

[NXL+07] Xiaochuan Ni, Gui-Rong Xue, Xiao Ling, Yong Yu, and Qiang Yang. Exploring in the weblog space by detecting informative and affective articles. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 281–290, New York, NY, USA, 2007. ACM.

[ON06] Jon Oberlander and Scott 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.

[PE05] Ana-Maria Popescu and Oren 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 339–346. Association for Computational Linguistics, 2005.

[Pew07] Pew Internet and American Life Project. Teens and social media, 2007.

[Pew08] Pew Internet and American Life Project. Internet activities, 2008.

[QC06] Yan Qi and K. Selçuk Candan. Cuts: Curvature-based development pattern analysis and segmentation for blogs and other text streams. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 1–10, New York, NY, USA, 2006. ACM.

[Sch07] Jan Schmidt. Blogging practices: An analytical framework. *Journal of Computer-Mediated Communication*, 12(4), 2007.

[SG06] Stephanie Smale and Saul 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.

[SOHB06] S. Sood, S. Owsley, K. Hammond, and L. Birnbaum. Reasoning Through Search: A Novel Approach to Sentiment Classification. *submitted to EMNLP, July*, 2006.

[Soo07] Sara Owsley Sood. *Compelling Computation: Strategies for Mining the Interesting*. PhD thesis, Northwestern University, Evanston, IL, USA, 2007. Adviser-Kristian Hammond.

[TC06] Chun-Yuan Teng and Hsin-Hsi Chen. Detection of bloggers' interests: Using textual, temporal, and interactive features. In *WI '06: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 366–369, Washington, DC, USA, 2006. IEEE Computer Society.

[Wil07] Matthew 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.

[WKY96] Jacqueline W. T. Wong, W. K. Kan, and Gilbert Young. Action: automatic classification for full-text documents. *SIGIR Forum*, 30(1):26–41, 1996.

[Wri06] Judith Wright. *The Soft Addiction Solution*. Tarcher, 2006.

[YNBN03] Jeonghee Yi, Tetsuya Nasukawa, Razvan Bunescu, and Wayne 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.