An Experiment on Semantic Emotional Evaluation of Chats

Concepción Bueno, Juan Antonio Rojo, Pilar Rodriguez
Escuela Politécnica Superior, Universidad Autónoma de Madrid,
Francisco Tomás y Valiente, 11
28049 Madrid, Spain
e-mail: diazmunio@hotmail.com, joan.rojo.83@gmail.com, pilar.rodriguez@uam.es

***Abstract*—Electronic conversations always contain an emotional charge. Being able to evaluate such emotional charge is an interesting challenge, and valuable conclusions can be obtained if that process is performed automatically. In this paper, we present a Semantic Emotional Evaluator for Chats, named Chat-SEE, that has been used for evaluating the emotions in a chat conversation. The results obtained are quite promising.**

Keywords- semantic emotional evaluation; cooperative work; chat semantic

I. Motivation

Electronic conversations, as well as any other kinds of conversations, always contain an emotional charge. Being able to evaluate such emotional charge is an interesting challenge, and valuable conclusions can be obtained if that process is performed automatically.

For those reasons, we planned to explore the possibility of designing and implementing an emotional evaluator that allows the measurement of the emotional content within a conversation. The emotional evaluation performed allowed us to research about the evolution of the participant emotions through the conversation. In the experiment carried out, three persons were asked to accomplish a cooperative task only making use of a standard online chat. In addition, they did not know who the other participants were.

In that context, our study aimed at proving that emotions can be measured and, also, that they present some relations among each other. Moreover, we aimed at presenting the results obtained in a visual and clear way.

In this paper, we present a Semantic Emotional Evaluator for Chats, named Chat-SEE, that has been used to evaluate the emotions in a chat conversation. The rest of this work is organized as follows. Section II briefly reviews the state-of-the-art on emotional evaluation. Following, Sections III and IV describe the experiment carried out and Chat-SEE evaluation steps, respectively. We present the results in Section V and the conclusions and future work in Section VI.

II. Related Work

Nowadays, emotional analysis is an outstanding research area that starts offering very interesting results. There are different studies, systems and applications available, which deal with emotion evaluation from diverse points of view. Some of them are based on voice spectrum and stress [1][2] or on gesture and expression analysis [3][4][5], while others try to conclude about the emotion charge of texts. Within this last group, some works have centered in analyzing individual short texts [6][7][8], while others have been applied to cooperative texts involving several users [9][10][11]. Though sharing similar goals, their approaches and final results are different.

For instance, [6] presents an approach to emotion analysis of new headlines. It proposes and evaluates several methods to identify an emotion in text. The emotions gathered are joy, anger, disgust, fear, sadness and surprise, which are used to classify the headlines accordingly. Also focused on headlines, the Headline Analyzer online application [7] aims at measuring the impact of short texts on potential readers, the so called emotional marketing value. In [7], the dimensions taken into account are: intellectuality, empathy and spirituality. Another market oriented application can be found in [8]: SAS Sentiment Analysis. That application analyzes digital content in order to understand customers' opinions. Positive and negative sentiments are inferred.

Regarding works that have centered on collaborative texts, [9] presents a study performed at HP Labs that demonstrates how important is to extract the emotions automatically from text in social media, and how it can be useful to forecast the impact of some topic. In particular they use tweets related to a movie to forecast box office revenues for movies. The emotions extracted from tweets are positive, negative and neutral. Once they extract the emotions from tweets and, applying different formulas, they obtain the positive or negative impact of a movie and, consequently, the higher or lower box offices revenues.

Also, a quite interesting work is presented in [10]. In that study they present a system to extract sentiment from text. It uses an annotated dictionary where a measurement of polarity, strength, intensification and negation are assigned to words. Different dictionaries are used with different results; it demonstrated the vital importance of the dictionary used. It is a content independent based system that has performed well on blog postings and video games reviews without any training process.

Finally, other interesting system is Text Tone [11]. Text Tone allows users to tag emotions in the text introduced in online textual communication, so people can easily understand the meaning of a conversation. It is useful when users try to express an emotion that can be ambiguous or to emphasize certain emotion. However, Text Tone does not analyze the text introduced by the user. There, users decide on their own emotion charge.

III. The Experiment

In this work, we took Gmail conversation chats as starting point. Each chat took place among exactly three people, being all of them students enrolled in the Master on Computer and Telecommunication Engineering at the Universidad Autónoma de Madrid. There were 6 groups, that is, 18 persons involved. Spanish was the language used.

Each group was asked to carry out a collaborative task during two hours. The activity consisted in trying to reconstruct a previously fragmented play script. There were some basic rules: they were not allowed to identify themselves and they did not have to delete anything they do. With this intention, each member of a group was provided with an e-mail address, its password and the e-mail address of his/her partners. Each team member had in the inbox of his/her e-mail a document with some characters of the play and some utterances. The whole play consisted of four characters and forty two utterances. Each group was required to give a joint solution to the activity. In order to do that, they had to gather all the information, attribute utterances to the characters, and chronologically arrange them. The process was unsupervised.

It is not surprising that the final chats became something funny and a little bit chaotic. When reading those chats, it seemed that people had had different attitudes when facing the proposed task, enjoying (or not) themselves during the process.

Then, we tried to determine whether the emotions in the conversation could be somehow evaluated. In that sense, we first had to decide which emotions we would focus on. Finally, we decided to make use of the classification ºproposed in [12]. In that work, four basic emotions are identified: joy, anger, fear and sadness. Authors state that those four basic emotions are directly related to the so named “fundamental challenges” such as danger (leading to fear), separation from positive conditions, including inadequate self-efficiency (leading to sadness), frustration of expectancies and registration of inhibitions (leading to anger) or self-efficiency and social acceptance (producing joy).

Though many other classifications of emotions can be found, as in the systems mentioned in Section II, we thought that the abovementioned classification fits perfectly for the experiment. The emotional meaning attributed to joy, anger, fear and sadness in Chat-SEE environment is briefly explained in next section.

IV. Chat-SEE

We have implemented Chat-SEE in a modular way, and based on three different modules: the dictionary, the tagger and the graph generator.

In addition, conversations are first converted to an XML format, so the rest of the process can be afforded easily. Programming language was Python, making use of the Natural Language Toolkit (NLTK) offered [13] [14].

In the rest of this section, the three modules are described. The examples and graphs presented are taken from a couple of chats, which correspond to groups A and B. Members of group A are identified as Huey, Dewey and Lowie, and members of group B as Kate, Jack and James.

A. Dictionary

Firstly, we created a dictionary based on the words that we had found in the chat to be used in this experiment. At this stage, no preprocessing, stemming or other NLP techniques were used. That decision was taken because of the characteristics of the texts: lots of misspellings and abbreviations.

In that process, not all the words presented in the conversations were tagged. The only words tagged were those that were supposed to have an emotion charge in the chat context.

The chat texts were initially XML formatted, so human judges could easily assign values to the different emotional dimensions chosen. More of a hundred of words were tagged, apart from some commonly used emoticons, what represented about 6-7% of the total words in the chats. In average, the total number of words in the chats was around 1500. For emotion quantification, it was decided to use a range between 0 and 3 (0 minimum and 3 maximum).

Regarding the meaning attributed to joy, anger, fear and sadness in Chat-SEE environment, it slightly differs from the meaning used in [12], being adapted to what a single word can express in terms of emotions. So, “anger” was also supposed to express a kind of criticism, as in the word “no”, whose entry is:

<**word** **token**="no" **joy**="0" **anger**="2" **fear**="1" **sadness**=”1” />

Also in that entry, a value of 1 for fear and sadness is attributed.

Other entries are simpler, like “ok”, that only seems to express some kind of “approval”, which is associated to joy:

<**word** **token**="ok" **joy**="2" **anger**="0" **fear**="0" **sadness**=”0” />

Some cross-checking of the emotion assignment was done in order to detect judge dependencies but most of the assignments were identical, or almost identical.

B. Tagger

The second stage in the emotional evaluator development is the creation of a parser-tagger. The main function of this parser-tagger is to isolate and to emotionally classify each word in an XML file. As was mentioned above, the creation of a structured file makes easier the measurement process for each user intervention. Also, NLTK Python module was used at this stage in order to carry out the process of word extraction and detection. Words are searched in the dictionary and, each utterance emotions are measured and assigned to each user intervention. The global emotions correspond to the sum of all the word emotions that appeared in the utterance.

Once we had all the emotional scores associated to each utterance, then we create a new XML file with the utterance scores. This file is the input for the last module of Chat-SEE: the Graph Generator. Apart from the emotion information per utterance, such file also includes a time stamp that makes possible to determine when each utterance had taken place.

Following, it is included part of the chat at this stage. It corresponds to the conversation taking place at the 27th minute within group A, among Huey, Dewey and Lowie. As it can be observed, during that minute Huey and Lowie make two contributions, whereas Dewey makes just one. In the text, **j**, **a**, **f** and **s** stand for joy, anger, fear and sadness, respectively.

```
<time id="16:51">
    <user id=“Huey”>
        <utterance>
            <word j=”1” a=”2” f=”1” s="1" token="ya"/>
            <word j=”0” a=”1” f=”0” s="0" token="veo"/>
        </utterance>
        <utterance>
            <word j=”0” a=”1” f=0” s="0" token="hay"/>
        </utterance>
    </user>
    <user id="Louie">
        <utterance>
            <word j=”0” a=”1” f=”0” s="0" token="ver"/>
        </utterance>
    </user>
        <user id=“Dewey”>
        <utterance>
            <word j=”2” a=”0” f=”0” s="0" token="vale"/>
            <word j=”0” a=”2” f=”1” s="1" token="no"/>
            <word j=”0” a=”2” f=”0” s="1" token="pero"/>
            <word j=”0” a=”1” f=”1” s="1"
token="entender"/>
            <word j=”0” a=”1” f=”1” s="1"
token="orden"/>
        </utterance>
    </user>
    <user id=“Huey”>
        <utterance>
        </utterance>
    </user>
    <user id=“Louie">
        <utterance>
        </utterance>
        <utterance>
        </utterance>
</time>
```

As can be seen, during the 27th minute Huey contributed twice to the chat, but only his first contribution had any emotion charge.

C. Graph Generator

Finally, Chat-SEE generates visual representations of the emotional evaluations by making use of standard graph generators, like GNUPLOT.

The Graph Generator of Chat-SEE works as follows. Firstly, it checks about the chat participants, and auxiliary files are generated for any of them, separately. Secondly, aggregation files are created for any of the utterances, by adding the emotion charges corresponding to each word. For Huey’s 27th minute, the result is:

```
<time id="16:51">
    <user id=“Huey”>
        <utterance w=”2” j=”1” a=”3” f=”1” s="1">
        </utterance>
        <utterance w=”1” j=”0” a=”1” f=”0” s="0">
        </utterance>
    </user>
    <user id=“Huey”>
        <utterance w=”0” j=”0” a=”0” f=”0” s="0">
        </utterance>
    </user>
</time>
```

In the former text, **w** indicates the number of words with emotion charge in each utterance, while the individual words have been eliminated.

Next, the resulting value assigned to each utterance is aggregated together with the values assigned to the rest of utterances that took place at that very minute. That value is divided among the number of contributions that had taken place at the same time. So, the final emotion media per user and minute are obtained. For Huey’s 27th minute, we obtain a emotion media of 0,5 joy, 2 anger, 0,5 fear and 0,5 sadness.

But, in this experiment, analyzing the evolution of the participant emotions was also a challenge. So, another kind of graphs was foreseen. In those graphs, the evolution of the participant emotion would be represented. Those graphs should smooth out the variation intensity of the participant emotions in the period under study.

For generating those smoothed-graphs, the emotion media previously calculated is divided by the number of instants (minutes in this case) since the beginning of the chat. Tables 1, 2 and 3 show all the emotion data for the members of group A (Huey, Dewey and Louie): utterances, contributions, emotion media and smoothed out emotion media at the 27th minute.

TABLE I. HUEY’S EMOTION CHARGE, 27TH MINUTE

		BASIC EMOTION VALUES			
CONTRIBUTION	**UTTERANCE**	**Joy**	**anger**	**fear**	**sadness**
contribution 1	utterance 1	1	3	1	1
	utterance 2	0	1	0	0
contribution 2	utterance 1	0	0	0	0
	EMOTIONS	0,5	2	0,5	0,5
	SMOOTHED EMOTIONS	0,8	0,57	0,17	0,36

TABLE II. DEWEY'S EMOTION CHARGE, 27TH MINUTE

		BASIC EMOTION VALUES			
CONTRIBUTION	**UTTERANCE**	**Joy**	**anger**	**fear**	**sadness**
contribution 1	utterance 1	2	6	3	4
	EMOTIONS	2	6	3	4
	SMOOTHED EMOTIONS	1,37	0,94	0,35	0,64

TABLE III. LOUIE'S EMOTION CHARGE, 27TH MINUTE

		BASIC EMOTION VALUES			
CONTRIBUTION	**UTTERANCE**	**Joy**	**anger**	**fear**	**sadness**
contribution 1	utterance 1	0	1	0	0
contribution 2	utterance 1	0	0	0	0
	utterance 2	0	0	0	0
	EMOTIONS	0	0,5	0	0
	SMOOTHED EMOTIONS	0,29	0,5	0,22	0,35

V. RESULTS

After Chat-SEE execution, three different kinds of graphs are obtained: instant emotion media per participant graph, smoothed out emotion evolution per participant graph and smoothed out chat evolution per emotion graph.

Figure 1 depicts Huey's emotion media during the 70 minutes that the experience lasted. In Figure 1, x-axis corresponds to moments (in minutes) and y-axis corresponds to the instant emotion intensity. In this kind of graphs, it is possible to detect when the emotion peaks took place at a glance. For example, in Figure 1 it is possible to observe that Huey's maximum "joy" happened a little bit after the 40th minute.

Regarding the second kind of graphs, which represent the smoothed out emotion evolution per participant, an example is presented in Figure 2, where x-axis corresponds to moments (in minutes) and y-axis corresponds now to the smoothed out emotion intensity. There, Huey's smoothed out emotion evolution is presented. Firstly, Huey seems to be quite expressive. Moreover, his "joy" line is high, and it surpasses the rest of his emotions. One possible interpretation is that Huey was motivated at accomplishing the proposed task and enjoyed himself while performing it.

Also, Huey "anger" line is not so relevant. It might be because, though he enjoyed himself, he did not take a leadership role.

Finally, Figures 3 to 6 represent the smoothed out emotions of the above mentioned groups, A and B, along the time. Both groups took part in the same experiment, as described in Section III. Those graphs represent the smoothed out chat evolution per emotion graph for both groups.

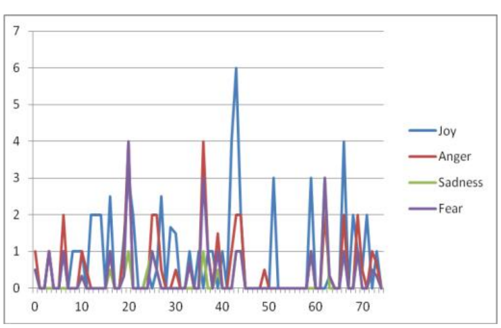

Figure 1. Huey's instant emotion media. A "Joy" peak takes place around the 40th minute of the experiment.

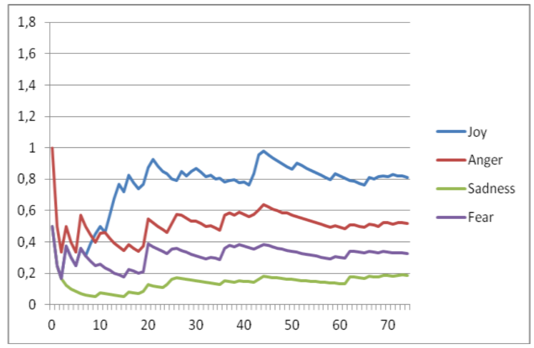

Figure 2. Huey's smoothed out emotion evolution during the experience.

In those figures, both Dewey and Jack seem to be the most expressive member of their groups, group A and B, respectively. As can be observed, both of them have the highest lines of their respective group in all the four emotions considered.

In addition, it is interesting to observe that the levels of "joy" and "anger" of both groups, A and B, is higher than their levels of "fear" and "sadness". From that, we could infer that the participants of both groups felt fine, they did not feel under pressure and, somehow, enjoyed themselves.

VI. CONCLUSION AND FUTURE WORK

In this work, we aimed at presenting an experiment on semantic emotional evaluation of chats. There are already some previous works in semantic emotional evaluation, as the ones mentioned in Section II, but they differ from Chat-SEE goals in several senses.

On one hand, Chat-SEE makes use of a different emotion classification, which, though taken from the psychological research area [12], has been re-interpreted in order to be used in our chat environment.

On the other hand, we were mainly interested in the emotion evolution from a relative point of view; that is: the emotion evolution among members of a group which were faced to work out a task collaboratively. So, we put more emphasis on the conclusions that could be derived within each group, rather than on the individual scores.

In that sense, Chat-SEE has obtained interesting results, because we have been able to measure how emotions evolve in an electronic conversation, being able to somehow "quantify" how they evolve. Moreover, Chat-SEE seems to be able to identify some kind of leadership role within conversations, as could be the case with Dewey and Jack. Exploring that possibility also is part of our future work.

There are some other challenges we face after this experiment.

Firstly, it is clear that the emotional dictionary used becomes a key module in the process, given that a bad emotional dictionary would clearly bias the final results. In

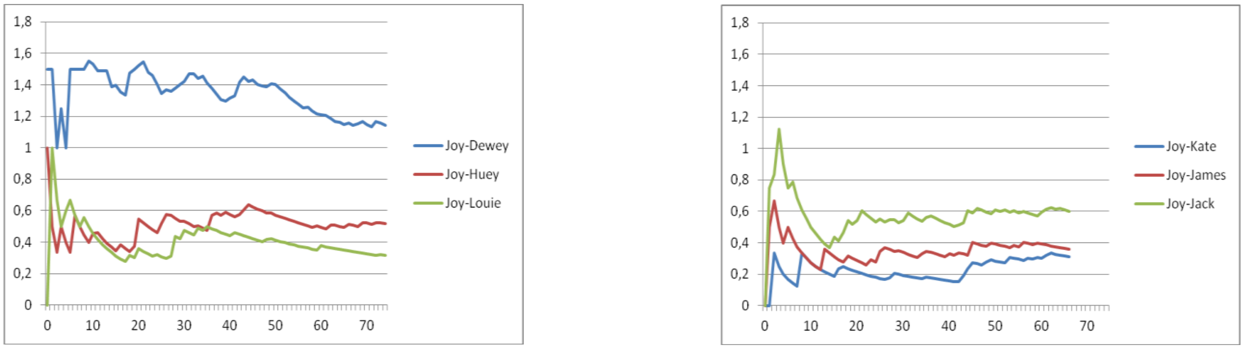

Figure 3. “Joy” representation for groups A and B.

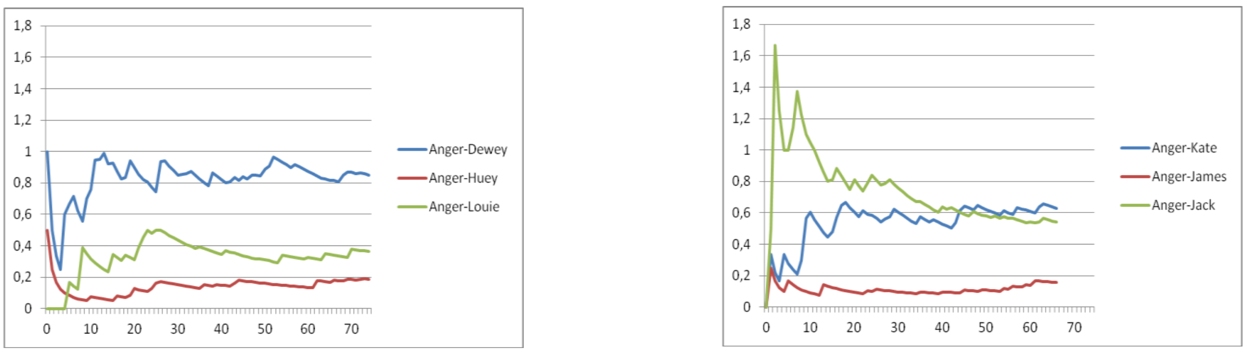

Figure 4. “Anger” representation for groups A and B.

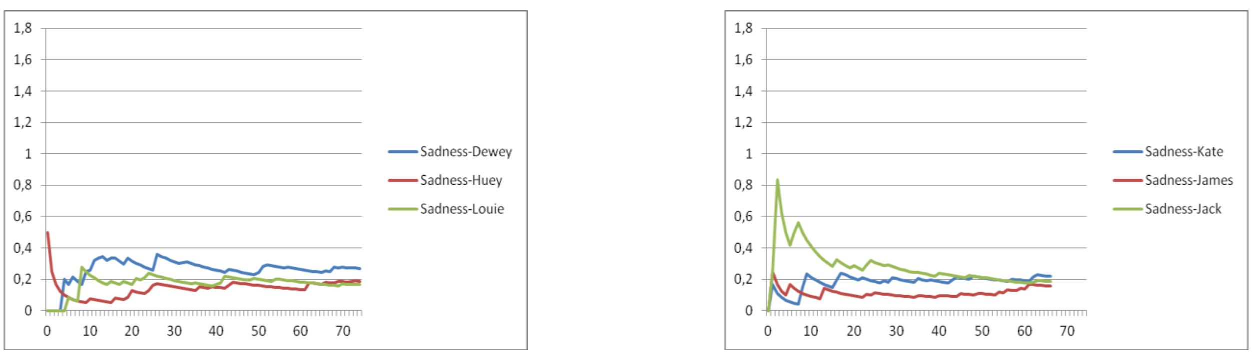

Figure 5. “Sadnes” representation for groups A and B.

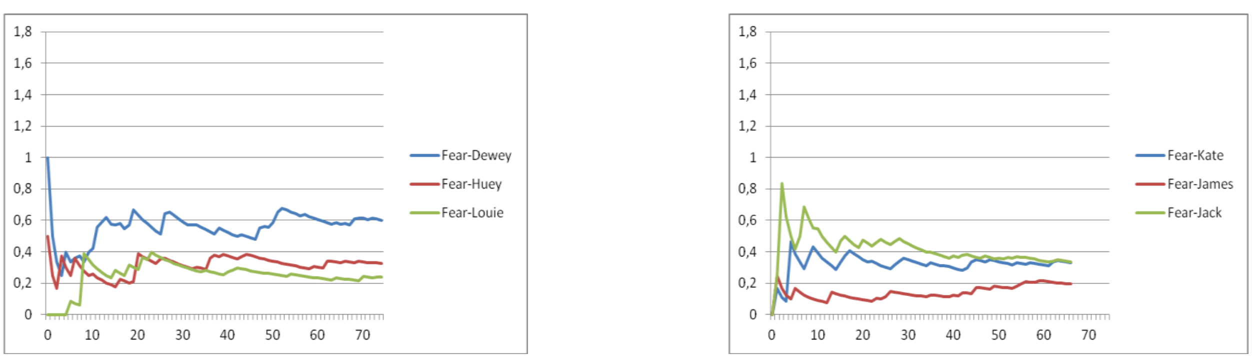

Figure 6. “Fear” representation for groups A and B.

that sense, we are aimed at improving the dictionary in two ways:

a) by including some kind of natural language preprocessing before the semantic emotion annotation,

b) by stablishing a judge protocol that would validate the semantic emotion asignment.

Moreover, the accumulation algorithm used has also become as a key module. We could modify our algorithm in several ways: media per paragraph, etc. Also, we could modify different parameters, as well as the weight given to them, by assigning different weight to the emotional dimensions depending on the chat subject. A comparative human analysis of the emotions of the chat is foreseen, in order to evaluate the correctness of the evaluation.

Finally, we plan to develop a graph zoom to be used for zooming instant peaks, and implement an online evaluator integrated in a chat tool. That online evaluator would let supervisors to react if some situations are identified.

ACKNOWLEDGMENT

This work has been partially supported by the Spanish Ministry of Science and Education (TIN2010-17344) and Comunidad Autonoma de Madrid (S2009/TIC-1650).

REFERENCES

[1] Ex-Sense, <http://www.ex-sense.com/proversion.html> 09.26.2011

[2] Á. Rodríguez Bravo et al., “Modelización acústica de la expresión emocional en el español.” Procesamiento del lenguaje natural, nº. 25, Sept. 1999, pp. 159-166.

[3] Glad or sad, <http://www.gladorsad.com/en/>09.26.2011

[4] MultimediaN, Emotional Analyzer, <http://www.multimedian.nl/en/demo_emotional_analyzer.php>09.26.2011

[5] A. Friberg, “A fuzzy analyzer of emotional expression in music performance and body motion”, Proc. Music and Music Science, 2004, pp. 1-13.

[6] C. Strappavana and R. Mihalcea, “Learning to identify emotions in text,” Proc. 23th ACM Symposium on Applied Computing (SAC’08), ACM Press, 2008, pp. 1556-1560, , doi:10.1145/1363686.1364052.

[7] Headline Analyzer, <http://www.aminstitute.com/headline/> 09.26.2011

[8] SAS ® Sentiment analysis, <http://www.sas.com/text-analytics/sentiment-analysis/> 09.26.2011

[9] S. Asur and B. A. Huberman, “Predicting the Future with Social Media”, Proc. WI-IAT’10, 2010 IEEE/WIC/ACM I nternational Conference on Web Intelligence and Intelligent Agent Technology, IEEE Computer Society, 2010, doi: 10.1109/WI-IAT.2010.63

[10] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, “Lexicon-Based Methods for Sentiment Analysis”, Computational Linguistics, vol. 37:2, Jun. 2011, pp. 267-307, doi:10.1162/COLI_a_00049

[11] A. Kalra and K. Karahalios, “TextTone: expressing emotion through text.” Interact 2005, LNCS 3585, Springer, 2005, pp. 966-969, doi: 10.1007/11555261_81

[12] A. Zinck and A. Newen, “Classifying emotion: a developmental account.” Synthese, vol 161:1, Jan. 2008, pp. 1-25, doi: 10.1007/s11229-006-9149-2

[13] D. Mertz, “Charming Python: Get Started with the Natural Language Toolkit.”, 2004, <http://www.ibm.com/developerworks/linux/library/l-cpnltk.html > 09.20.2011

[14] Natural language toolkit, <http://www.nltk.org/> 09.20.2011

