

## **SocialMood: A visualization tool to measure the mood of students in social networks**

**Guilherme de O. Amorim<sup>1</sup>, Rodolfo B. Moraes<sup>2</sup>, Roberto Y. da S. Franco<sup>3</sup>, Ricardo A. Afonso<sup>4</sup>**

<sup>1, 2, 3, 4</sup>Universidade Federal de Alagoas – Arapiraca, AL – Brazil

{goa2, rbm3, rysf}@cin.ufpe.br, afonso055@gmail.com

***Abstract.** Based on the arena of social networks (SN), this study, through of a visualization tool aims to identify trends mood among undergraduate students. Combining the methodology Self-Assessment Manikin (SAM), which originated in the field of Psychology, the system filters the content provided on the Web and isolates certain words, establishing a range of values as perceived positive, negative or neutral. A data base summarizing the results, assisting in the construction and visualization of behavioral profiles generic, so we have a guideline for the visualization of information about SN through information visualization (IV) tool.*

### **1. Introduction**

Ellison (2007) said that social networking sites (SNSs) such as Friendster, MySpace, Facebook, Orkut, LinkedIn, and myriad have attracted millions of users. They keep saying that many of them have integrated SNSs into their daily lives to communicate, send and receive information in many ways.

As for Krebs (2008) the emergence of SNSs such as Facebook and MySpace has revolutionized the availability of social and demographic data, which has in turn had a significant impact on the study of SN. For analysis of this large quantity of information we have Sentiment Analysis (SA), where Liu (2010) defines it like the computational study of opinions, sentiments and emotions expressed in text. Thus, IV aims to assist users in exploring, managing, and understanding the ever-growing amount of digital information [Toker 2013]. In this way, the goal of this paper is facilitate the visualization of information about change in the mood of Brazilian undergraduate students for analyze what generate discomfort between them.

The paper is organized as it follows: in section II, the state-of-the-art of IV. In section III, Affective Norms explains words' normalization. Section IV, SocialMood's implementation. Section V, system's evaluations. Finally, conclusions and perspectives in section VI.

### **2. Information Visualization (IV)**

According Plaisant (2003), the area of IV, aims to design compact graphical representations and software interfaces that allow the user manipulation of a large number of items of visual data, possibly extracted from large databases. For this work we use PRISMA, a desktop tool based on multiple coordinated views, which has three main techniques: data dispersion, tree map and parallel coordinates. Additionally, PRISMA offers portability, usability and extensibility [Godinho 2007].

#### **2.1. Process of Visualization**

Normally, the process of IV is part of a process of exploratory data analysis, knowledge discovery, or visual analysis. The quality of this analysis is dependent on the quality of the data, with elimination of errors or noise data. Done that, a visualization tool contributes to

various user profiles: experienced, beginners, with more or less skill in the problem domain, until they feel comfortable with that is shown.

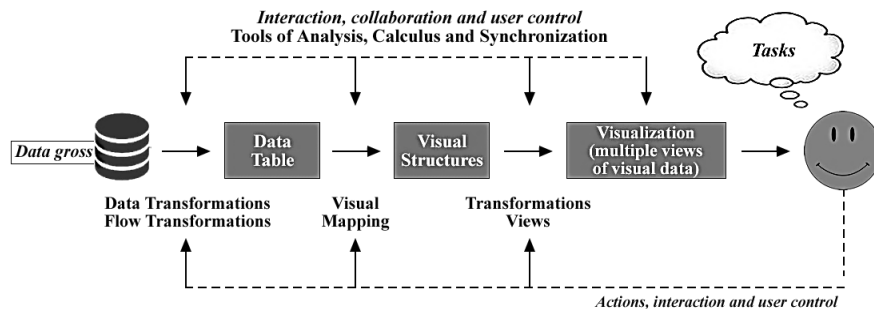


Figure 1. Visualization pipeline according to Ward, Grinstein and Keim (2010).

### 3. Affective Norms

In Bradley and Lang (1999), the Affective Norms propose to normalize human emotions assigning them a numerical value through an object of study. The Affective Norms for English Words (ANEW), for instance, classifies 1,040 words in English in terms of valence (ranging from pleasant to unpleasant) and arousal (ranging from calm to excited).

A common theme for affective norms analysis is classifying words using the SAM. This way, Stein (2009) generated 44 lists of semantically related Brazilian-Portuguese words classified with valence value, in all 660 words. Oliveira (2013) used a similar method to generate the valence values of a pool of 908 words. Finally, Kristensen (2011) evaluated the valence rating of 1,046 words in Brazilian-Portuguese.

#### 3.1. Affective Norms Classification

SocialMood worked with a pool of Brazilian-Portuguese words similar to the ANEW – the Brazilian norms for the Affective Norms for English Words [Kristensen 2011], which translated and adapted the ANEW to Brazilian-Portuguese (generating the ANEW-Br).

The ANEW-Br study was conducted to collect emotionality measures for a set of 1,040 Brazilian Portuguese words. A group of 755 Brazilian undergraduate students rated these 1,040 words using a paper-and-pencil version of the valence and the arousal scales of the SAM, each scale ranging from 1 to 9. Soon after, data were computed and generated a table with valence and arousal norms for the classified words.

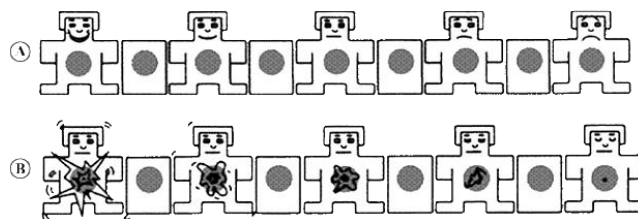


Figure 2. Scales Self-Assessment Manikin (SAM) of valence (A) and arouse (B).

For our work, the arousal measure is not considered. SocialMood uses the valence data of words from a dataset composed of the ANEW-Br pool of words and pools from others similar papers.

### 4. SocialMood

SocialMood aims to measure the students' mood in posts on SN (in this paper, Facebook) and show this information though a visualization tool. It uses the Sentiment Analysis (SA) area, as

a basis for work. SA, also known as opinion mining, grows with the need to find different opinions hidden in long forum posts and blogs, e.g. This way, Godbole, Namrata, Srinivasaiah, and Skiena (2007) mentioned it is a challenging natural language processing or text mining problem. SA divided into four major steps [Liu 2005]: (1) detect if a text is subjective or objective; (2) detect the characteristics of the object under examination; (3) classify the text into a positive, negative or neutral; (4) view the results through various forms of visualization. Steps (3) and (4) are our focus.

#### 4.1. Architecture

The system proposes a modular architecture. It is divided in three principal components: Dataset Manager, a data base with different data types; SocialMood, the application that infers valence scores for posts; and PRISMA, visualization component.

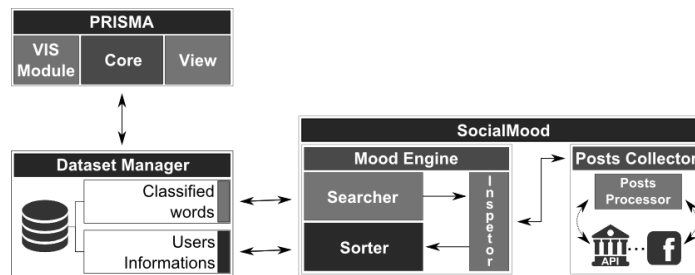


Figure 3. SocialMood Architecture.

The SocialMood component consists of two modules: Mood Engine, which is responsible for classifying; Post Collector, which returns the user's Facebook posts in analysis. Mood Engine works with the Searcher, which searches in the Dataset Manager words that are found in the posts; Inspector receives the words of the database and posts of the Post Collector, and search the words classified within the posts; Sorter performs the calculation shown in section 4.2 and stores the results in the Dataset Manager User Information; PRISMA, together with DasetManager, infers easy viewing of dates, and it got a better analysis of the data.

#### 4.2. Key Features

SocialMood tries to represent the profile of mood changes. A table in our database gathers words classified by SAM, where that are fetched one by one in Facebook posts. After, the system sums up all values of valence of words in those texts and divides it by the total number of occurrence of words. Consider an excerpt from the lyrics of Tom Jobim, The Girl from Ipanema:

O  mundo  inteirinho se enche de graça / E fica mais  lindo  / Por causa do  amor

*The whole  world  is filled with grace / And it gets more  beautiful  / Because of  love*

We found three instances in database: mundo (world), lindo (beautiful), amor (love). Valence values are respectively: 6.17, 8.17, 8.75. Sum the values of the valences of words and divided by the amount of occurrence of words in the text, which in this example was 3:

$$\frac{6.17 + 8.17 + 8.75}{3} \approx 7.69$$

Thus, the valence of our post is 7.69, which according to Kristensen (2011), is valence for positive. So, our excerpt Music Tom Jobim is positive. So, this process applied to the user's post, we can infer whether the post is positive, negative or neutral. With this we can infer user's mood analyzing their posts.

### 4.3. Prototype and Features

Preliminary analysis of SocialMood presented some problems. Initially, the database of classified words SocialMood is still small (with only 2614 words) compared to dictionaries of Portuguese as the Aurélio dictionary online, that cataloged 435,000 words. It gives us less precision to classify posts coming from Facebook. In addition, problems such as inferring the correct semantics of sentences, like ironies, have not yet been solved.

## 5. Evaluations

The validation process was based on the comparison of the results produced by the tool SocialMood and compared with The Positive and Negative Affect Schedule (PANAS) [Watson and Clark 1999].

### 5.1. Hypotheses

Rocha and Baranauskas (2003) emphasize the importance of clearly defining the objectives before the test definition, since these must be performed in order to highlight the usability criteria to be measured. At first, we consider accuracy (accuracy, number of hits): the hit rate is lower SocialMood tool axis relative publications PANAS negative mood. This is possibly due to the information not these correspond to the reality of users. Second hypothesis is about the hit rate of the tool is consistent SocialMood the shaft PANAS publications regarding positive mood. We believe that users should show greater reality to positive situations.

### 5.2. Results

The graph in Figure 4 shows that use of neutral words is linear. Negative words, had a peak on Monday, however, declines. Positive words are the opposite of the negative: starts with little occurrence and increases. From the graph we can analyze the behavior of the student through what he posts.

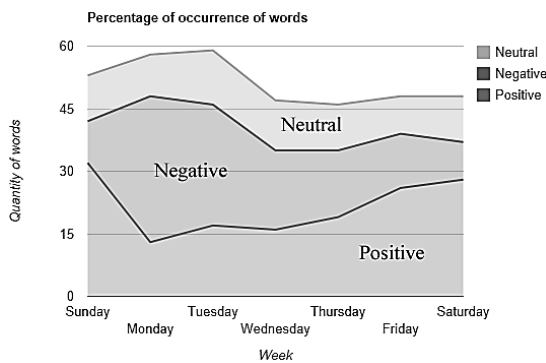


Figure 4. Number of words per day.

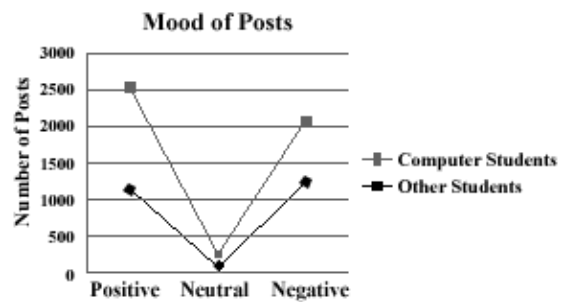


Figure 5. Mood of Posts.

The tool was able to identify the behavior of computing students compared with others (Figure 5). This is due to the possibility of computing students have greater ability to interact man x computer.

## 6. Conclusions and perspectives

SocialMood proved efficient to draw a profile of humor undergraduates. Thus, we can monitor the student's behavior through his Facebook posts observing changes in their behavior. This way, university team can interfere positively on the student, helping to solve potential problems. We will expand the coverage of the software, and do more tests like performance and usability. For more accurate post classification, still necessary to find a method to solve

problems in context in sentences. Like future work we will develop a web application for Facebook that implements the SAM, for classify more words. In addition, we will implement a solution for context problem, like irony.

## References

- Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (pp. 1-45). Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.
- Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.
- Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. *ICWSM*, 7.
- Godinho, P., Meiguins, B., Carmo, C., Garcia, M., Almeida, L., & Lourenço, R. (2007). PRISMA - A Multidimensional IV Tool using Multiple Coordinated Views. 11th International Conference on Information Visualization, (pp. 23-32). Zurich.
- Krebs, V. (2008). Social network analysis. Retrieved on April, 26, 2008.
- Kristensen, C. H., Gomes, C. F. D. A., Justo, A. R., & Vieira, K. (2011). Brazilian norms for the Affective Norms for English Words. *Trends in Psychiatry and Psychotherapy*, 33(3), 135-146.
- Liu, B., Hu, M., & Cheng, J. (2005, May). Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web* (pp. 342-351). ACM.
- Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2, 568.
- Oliveira, N. R. D., Janczura, G. A., & Castilho, G. M. D. (2013). Norms of arousal and valence for 908 Portuguese words. *Psicologia: Teoria e Pesquisa*, 29(2), 185-200.
- Plaisant, C., Chintalapani, G., Lukehart, C., Schiro, D., & Ryan, J. (2003, October). Using visualization tools to gain insight into your data. In *SPE Annual Technical Conference and Exhibition*.
- Rocha, H. V., & Baranauskas, M. C. (2003). *Design e Avaliação de Interfaces Humano-Computador*. Campinas: Instituto de Computação - Universidade Estadual de Campinas.
- Stein, L. M., & Gomes, C. F. D. A. (2009). Brazilian norms for word lists: semantic association, concreteness, word frequency, and emotionality. *Psicologia: Teoria e Pesquisa*, 25(4), 537-546.
- Toker, D., Conati, C., Steichen, B., & Carenini, G. (2013, April). Individual user characteristics and IV: Connecting the dots through eye tracking. In *Proceedings of the SIGCHI Conference Human Factors in Computing Systems* (pp. 295-304). ACM.
- Ward, M. O., Grinstein, G., & Keim, D. (2010). *Interactive Data Visualization: Foundations, Techniques, and Applications*.
- Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*.