Identifying Emotions in Short Texts for Brazilian Portuguese

Barbara Martinazzo, Mariza Miola Dosciatti, Emerson Cabrera Paraiso

Graduate Program on Informatics – PPGIa, Pontifícia Universidade Católica do Paraná – PUCPR, Curitiba – PR – Brazil
{martinazzo, mariza.dosciatti, paraiso}@ppgia.pucpr.br

Abstract. The automatic detection of emotions in texts has presented significant results in several and different situations. In this paper, we present an approach to identify automatically emotions in short texts written in Brazilian Portuguese. Each text is processed using an algorithm based on the Latent Semantic Analysis theory. Experimentations have shown that this method can identify the correct emotions in short texts in about 70% of the cases.

Keywords: Emotion identification; LSA; Short Texts; Brazilian Portuguese

1 Introduction

Recent advances in texts analysis lead to the emergence of a new area responsible for the recognition of subjective aspects, such as opinions, feelings and emotions in texts. Research in this area refers to the development of methods to allow computational systems to be able to recognize and detect affective factors in texts. However, as it is a relatively new area, these methods are still under development and are, in its vast majority, only for the English language. Thus, we notice the need for adaptation of these methods to other languages, such as Brazilian Portuguese. The main goal of this research is to propose a method, based on the Latent Semantic Analysis, for emotion identification in texts for Brazilian Portuguese. In the context of this research, the texts that are been used are short news (headlines and a short description). Conversation interfaces are one for the several applications that could be improved with this kind of research.

The rest of the paper is organized as follows: section 2 gives a short overview on emotions identification in texts, section 3 presents our approach to identify emotions in short text for Brazilian Portuguese with some experimental results. Finally, we present some conclusions and indicate some perspectives for future work.

2 Identifying Emotions in Texts

This session presents the main concepts behind the process of identifying emotions in texts. We start by defining Emotions in this context.

adfa, p. 1, 2011.

2.1 Emotions

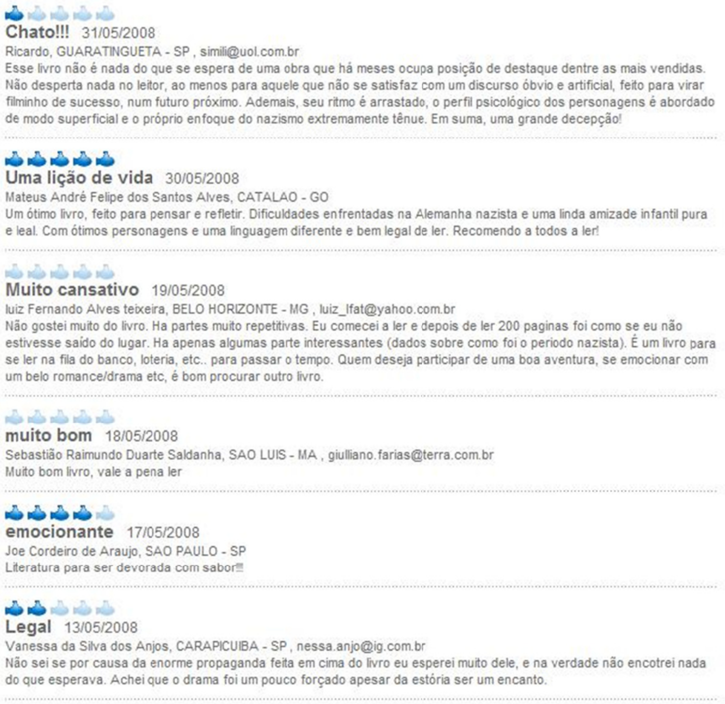
Fehr and Russell [1] say that unless they are asked for a definition, all people know what an emotion is. They even question whether emotions can be considered as psychological, mental, or behavioral events, and also if there are emotions which can be considered more "basic" than others. Although there is still no concrete definition about this description, we will assume that emotions are mental and psychological states, associated to a great variety of feelings, thoughts and behaviors. Gazzaniga and Heatherton [2] say that emotions have been studied in several areas of the human knowledge for a long time. According to Strongman [3], Plato and Aristotle were the first ones to concern the subject. Aristotle believed that the emotion is the most interesting side of human existence. On the other hand, Darwin, in his book "The Expression of the Emotions in Man and Animals", emphasized the main role of emotions in the evolutionary process of all living beings.

Currently, researches and studies concerning emotions are divided in several and different areas, but the one which will be discussed and used in this paper is the simplest of them, named Basic (or Pure) Emotions. This concept is related to the innate emotions shared among all cultures in the world and it was proposed in the 1970's by Paul Ekman and Wallace Friesen [4]. Since there is no agreement about how many and which these emotions are, the model proposed by Ekman and Friesen mentioned six: sadness, anger, happiness/joy, fear, disgust and surprise. According to this theory, Ekman and Friesen decided to throw an experiment in different countries around the world, where they asked people to identify emotional responses they saw in pictures and facial expressions shown. Throughout this experiment, they found out that the six emotions proposed in this study were easily interpreted in all countries where the experiment took place.

Emotions have been subject of research for many different areas, such as psychology and other sciences responsible for the study of human behavior. Recently, these studies have also been attracting the attention of researchers in Computer Science, especially regarding the interaction between human and machines [5]. Among the subjects and the ongoing researches, we can cite the automatic recognition of emotions in texts, which is known as Sentiment Analysis.

2.2 Emotions Identification in Texts

It is known that, besides information, texts may also carry the expression of opinion and emotional state from the author. Systems used for product classification, for instance, most of the time do not show only facts, but mostly personal opinions. Thus, sentimental analysis of revisions made by customers may help the system to, based on previous written evaluations, recommend or not some product to a new costumer who is searching for information about it. An example that can be cited is the classification of books and movies as positive or negative by ranking it according not only to the traditional ranking system, but also based on textual evaluation left by other people who have already read or watched the story [6]. This would be extremely useful to establish a comparative between the written opinion and the ranking given by the customer. However, in Fig. 1 it is possible to observe that not always the ranking and the written opinion are consistent. It is important to remember that, since our research concerns the Brazilian Portuguese, some texts shown in examples were written in Brazilian Portuguese.



●○○○○
Chato!!! 31/05/2008
Ricardo, GUARATINGUETA - SP , simili@uol.com.br
Esse livro não é nada do que se espera de uma obra que há meses ocupa posição de destaque dentre as mais vendidas. Não desperta nada no leitor, ao menos para aquele que não se satisfaz com um discurso óbvio e artificial, feito para virar filminho de sucesso, num futuro próximo. Ademais, seu ritmo é arrastado, o perfil psicológico dos personagens é abordado de modo superficial e o próprio enfoque do nazismo extremamente tênue. Em suma, uma grande decepção!

●●●●●
Uma lição de vida 30/05/2008
Mateus André Felipe dos Santos Alves, CATALAO - GO
Um ótimo livro, feito para pensar e refletir. Dificuldades enfrentadas na Alemanha nazista e uma linda amizade infantil pura e leal. Com ótimos personagens e uma linguagem diferente e bem legal de ler. Recomendo a todos a ler!

○○○○○
Muito cansativo 19/05/2008
luiz Fernando Alves teixeira, BELO HORIZONTE - MG , luiz_lfat@yahoo.com.br
Não gostei muito do livro. Ha partes muito repetitivas. Eu comecei a ler e depois de ler 200 paginas foi como se eu não estivesse saído do lugar. Ha apenas algumas parte interessantes (dados sobre como foi o periodo nazista). É um livro para se ler na fila do banco, loteria, etc.. para passar o tempo. Quem deseja participar de uma boa aventura, se emocionar com um belo romance/drama etc, é bom procurar outro livro.

○○○○○
muito bom 18/05/2008
Sebastião Raimundo Duarte Saldanha, SAO LUIS - MA , giulliano.farias@terra.com.br
Muito bom livro, vale a pena ler

●●●●○
emocionante 17/05/2008
Joe Cordeiro de Araujo, SAO PAULO - SP
Literatura para ser devorada com sabor!!!

●●○○○
Legal 13/05/2008
Vanessa da Silva dos Anjos, CARAPICUIBA - SP , nessa.anjo@ig.com.br
Não sei se por causa da enorme propaganda feita em cima do livro eu esperei muito dele, e na verdade não encotrei nada do que esperava. Achei que o drama foi um pouco forçado apesar da estória ser um encanto.

Fig. 1. A book classification blog system (in Brazilian Portuguese)

As an example, where we can read “muito bom” (which means “very good”), we notice that the written evaluation was positive, while the ranking was 0 (fourth evaluation). Also, we notice that where we read “muito cansativo” (meaning “very tiring”), we can see that the written evaluation shows the reader did not like the book, and also ranked it as 0 (third evaluation). Throughout this snapshot, we notice that users are not always concerned to both aspects (ranking and written opinion), and this is a very common scenario in any evaluative system, which means the integration between written and ranked opinion through automatic analysis would be useful and help to improve this scenario and make it more trustable.

Although such analysis seems to be simple, it is actually a complicated task. The simple act of searching for words related to emotions (such as “good” or “bad”, “sad” or “happy”) in order to classify a document, according to the frequency these words appear, is not sufficient. Consider the following example: “The protagonist tried to protect his good name”. This sentence contains the word “good”, but there is no relevant information that can make it be considered a subjective affirmation. Pang [6] and CUCS in [7] presented an approach in which they tried to obtain better results making the distinction between subjective and objective sentences, which will be discarded. In their approach, they submit the remaining sentences (subjective sentences) to a classifier. CUCS [7] says that the biggest problems they faced consist, mainly, in: defining whether a sentence is subjective or objective; determining if the sentiment present is positive or negative; and, finally, determining how much of that sentiment is present in the sentence.

The richness of human language allows us to express one piece of information in many different ways, what makes it difficult to find the correct emotional indicatives within a sentence. As an example, we cite the following three sentences below, extracted from CUCS [7]:

- This laptop is a great deal.
- A great deal of media attention surrounded the release of the new laptop model.
- If you think this laptop is a great deal, I´ve got a nice bridge for you to buy.

In all cases, we notice the existence of the expression "great deal", but the opinions found in the sentences are, respectively, positive, neutral and negative. In the first two sentences, "great deal" has different meaning, and in the third case, it is used with sarcasm, which is something easily found in subjective texts, such as blogs, forums, etc.

One of the techniques used to identify the relation between words in a text is known as Latent Semantic Analysis, and it is going to be presented in the next paragraphs.

2.3 Latent Semantic Analysis

Latent Semantic Analysis (LSA) is a statistical/mathematical method used to identify relations between words in texts [8]. From these connections, it is aimed to establish associations among terms present in a majority of texts from one group of documents [9]. Traditionally, it is not considered a Natural Language Processing method, since it does not make use of any sort of dictionary or human-made basis, semantic networks, grammatical analyzers and so on. Besides that, the only input data used is a collection of text and/or sentences [8]. However, LSA has showed consistent results in many applications, including documents categorization based on conceptual similarities [10].

The first step for LSA execution is to represent the texts/passages through a term-document matrix, where each line represents one word (or term) from all the words found and each column represents one of the texts/passages (or document) in the collection. This matrix identifies the occurrence of terms inside a group of documents. Thus, each cell contains the frequency a word appears in a document. Then, each cell is submitted to a preliminary calculation process where to each frequency of occurrence will be given a value. There will be two different values: one for the importance of the word within each document and another for the whole collection of analyzed documents. Finally, a theorem called Single Value Decomposition (SVD) is applied to the matrix to determine patterns and relationships among the terms found ([8] and [10]). Based on these patterns, the next step is to make the correct approximations in order to obtain the desired grouping or classifications.

LSA was used in our method, and will be presented on the following section.

3 Identifying Emotions in Short Texts for Brazilian Portuguese

This research main goal is to automatically identify emotions in short texts written in Brazilian Portuguese. In order to test our approach, we have been tested it with short news and tweets. In this paper we are going to work only with short news.

3.1 Identifying Emotions in Short News

For the purpose of this work, it was defined that "short news" is actually short texts composed by the news' headline and a short description. This short news has a limited size in number of words, being similar to each other in terms of size. The headline's objective is to attract the reader's attention and the short descriptions have a brief explanation about the headlines. These two pieces of texts bring a lot of sentimental information and help the reader to decide whether he/she will continue reading the whole article or not. This is an example, in English, of a headline and its short description: "Nearly 100 Missing after Russian Riverboat Sinks (*headline*): A boat filled with families on the Volga River sank, and about 100 people were missing hours later, feeding fears that the episode could be the country's worst such accident in recent history (*short description*)"[1].

The next paragraphs describe the whole process.

The Training Step. In order to better illustrate the training process, we shall take as example, the short news from Table 1 (texts in Brazilian Portuguese).

Table 1. Short News Examples

News	Short text
1	Presidente do TCE no RS deixa hospital após assalto: João Vargas foi esfaqueado na barriga e passa bem em São Sepé. Dois suspeitos foram presos pela polícia na cidade de Santa Maria.
2	Dois morrem e um fica ferido após carro cair uma altura de 10 metros: Acidente aconteceu neste sábado em Caxias do Sul. Veículo caiu com as rodas para cima em uma represa vazia.
3	PRF flagra menina de 12 anos dirigindo picape em rodovia gaúcha: Segundo os policiais, jovem estava acompanhada da mãe em São Borja. Carro foi retido e a mãe autuada por deixar um menor de 18 anos dirigir.
4	Sucuri de 8 metros é flagrada após comer uma capivara: Cobra virou atração para moradores de São José do Rio Claro (MT). Ela foi encontrada em um riacho perto de uma fazenda na cidade.

[1] Extracted from The New York Times: http://www.nytimes.com/2011/07/11/world/europe/11volga.html.

5	Perito particular questiona imagens de acidente com ex-deputado no PR: Família de vítima contratou profissional para fazer simulação da colisão. Segundo ele, alguns segundos do filme do acidente foram removidos.
6	Depois da guerra, Faixa de Gaza vira 'ilha à deriva': Quatro meses após ataque de Israel, território palestino segue pressionado. Mas região deve voltar a ser o foco do processo de paz no Oriente Médio.
7	Prédio desaba e fere ao menos 3 na Bélgica: Acidente aconteceu durante festa local de Ducasse de Doudou. Para bombeiros, pode haver feridos ou mortos nos escombros.

Before to start, it is important to highlight that we assume that the words will always have similar meanings and will always be inserted within similar contexts ([9] e [10]).

The first step is to submit the whole set of texts to a preprocessing task. In this part of the process, the following will be done: firstly, the file containing the texts will be read and all letters will be lowercased, then all special characters (such as punctuations and numbers) and stop words will be removed. Then, a stemmer will be applied to the remaining words. This is extremely important to reduce all words to their radicals and, this way, to keep similar words together in one group, as for example the words "war" and "warrior". These two words' radicals are the same, "war", and this step helps us to group these words together, making the whole process of LSA more trustable. The removal of stop words and the stemmer were implemented with the help of the Weka [11] tool, which already implements an algorithm to stop words removal and an extension of Snowball Stemmer [12]. These resources were chosen because of the easiness to integrate to the project and also because of the availability to Brazilian Portuguese (stemmer). In Table 2, one can see the result of the preprocessing task for the examples shown in Table 1.

Table 2. Texts From Table 1 After Preprocessing

News	Preprocessed Short Text
1	tce rs deix hospital assalt joã varg esfaqu barrig pass sep suspeit pres sant mar
2	morr fic fer carr cair altur metr acident acontec cax veícul caiu rod cim repres vaz
3	prf flagr menin dirig picap rodov gaúch polic jov acompanh mã borj carr ret mã autu deix menor dirig
4	sucur metr flagr com capiv cobr vir atraçã morador clar mt encontr riach pert fazend
5	perit particul question imagens acident ex deput pr famíl vítim contrat profissional simul colisã segund acident remov
6	guerr faix gaz vir ilha deriv ataqu israel territóri palestin seg pression volt foc paz orient médi
7	prédi desab fer bélgic acident acontec fest duc doud bombeir hav fer mort escombr

After the preprocessing process described above, two vectors are generated. The first one keeps all the short news (documents) already preprocessed. From this first vector, a second one is created, which contains all the words (terms) found in the first vector (set of documents), without repetition. Based on these two vectors, a term-document matrix is created and, initially, filled with zeros. Each row of this matrix corresponds to a word from the second vector, and each column to a document from the first vector. Thus, each cell from the table corresponds to the frequency of occurrences of each term in each document. After the term-document matrix is fully built, all terms that appear only once within the set of documents are ignored. In Table 3, the term-document matrix created for the example is shown, where D1, D2, D3,... represent the short news from Table 1. It is important to highlight that the terms observed below are the stemmed representation of the original words in Brazilian Portuguese. This process is crucial in order to keep similar terms in radical grouped together.

Table 3. Term-Document Matrix Example (Excerpt)

Term	D1	D2	D3	D4	D5	D6	D7
acident	0	1	0	0	2	0	1
acontec	0	1	0	0	0	0	1
carr	0	1	1	0	0	0	0
deix	1	0	1	0	0	0	0
dirig	0	0	2	0	0	0	0
fer	0	1	0	0	0	0	2
flagr	0	0	1	1	0	0	0
metr	0	1	0	1	0	0	0
mã	0	0	2	0	0	0	0
vir	0	0	0	1	0	1	0

From this matrix, a second procedure is started. This one also uses tools from Weka and consists of the application of a technique called Singular Value Decomposition (SVD). The SVD is used in order to find a dimensionally reduced representation of the term-document matrix, which will emphasize the strongest patterns and relationships among terms and/or documents, at the same time that it will discard noises and the weakest relationships and patterns. The execution of SVD will decompose the term-document matrix into three new ones. The first one is called *U* and gives us the coordinates of each term inside the space of terms and documents being analyzed.

Table 4 shows a small part of the matrix *U* created using the above described process. Each row represents one term from the term-document matrix and each column represents one different coordinate.

Table 4. The Matrix *U* (an excerpt)

Term	Coordinate 1	Coordinate 2	Coordinate 3	Coordinate 4	Coordinate n
acident	*0.3306*	*0.5410*	*0.2103*	*-0.7412*	*0.0633*
acontec	0.2182	0.3307	-0.0311	0.3146	-0.1315
carr	0.3534	-0.0282	-0.0377	0.0669	-0.5667
deix	0.2581	-0.1928	0.1013	0.0046	0.1637
dirig	0.4723	-0.3481	0.1445	0.0059	0.0755
fer	0.3191	0.5156	0.0476	0.5652	0.3414

flagr	0.2769	-0.1780	-0.3898	-0.0869	0.1502
metr	0.1580	0.1419	-0.5720	-0.0259	-0.4920
mã	0.4723	-0.3481	0.1445	0.0059	0.0755
vir	0.0446	-0.0043	-0.6481	-0.1395	0.4880

The second matrix is called V_T and, like matrix U, it will give us the coordinates of each document inside the same space. Table 5 shows a small part of the matrix V_T, where D1, D2, D3, ... represent documents and each row represents one different coordinate.

Table 5. The Matrix V_T (an excerpt)

D1	D2	D3	D4	D5	D6	Dn
0.0752	0.4022	0.8099	0.1398	0.1928	0.0130	0.3461
-0.0601	0.4679	-0.5584	-0.0126	0.3372	-0.0014	0.5931
0.0543	-0.2051	0.1348	-0.8625	0.2253	-0.3472	0.1470
0.0027	0.1072	0.0050	-0.1506	-0.8846	-0.0833	0.4200
0.1436	-0.6891	0.0430	0.1282	0.1110	0.4280	0.5392

Finally the third matrix, called S, allows us to estimate how many dimensions should be used in order to obtain the best results. This number, according to Grossman e Frieder [14], can also be estimated arbitrarily through experiments and comparisons. In the given example, we adopted 5 dimensions because the number of texts is small. However, for the experiments done with the algorithm, the best results were obtained with 50 dimensions.

The next step is responsible for defining the position of each group (emotion) within that same space created before throughout the SVD process. In order to find the most appropriate location for each group, we used six sets of words, one for each emotion, where each list would contain some words related to each correspondent emotion. These six sets of words were manually built and contain approximately 800 words (examples in Table 6).

Table 6. Some Words for Each Emotion

Emotion	*Alegria* *(joy)*	*Desgosto* *(disgust)*	*Medo* *(fear)*	*Raiva* *(anger)*	*Surpresa* *(surprise)*	*Tristeza* *(sadness)*
Examples	amor amizade brincadeira esperança engraçado	enjôo feio náusea nojo sujo	assombrado cruel medroso pânico terror	assassinar cólera destruir diabólico irritar	deslumbrar embasbacar fantástico pasmo susto	arrepender chorar derrota desamparo luto
Quantity	278	72	104	168	40	184

In order to define the location of the groups for each emotion, an algorithm analyzes each list of words and searches for words contained in the terms vector. When it is done, it is able to determine, using matrix *U*, the center of the group for each emotion. This center is obtained by calculating the midpoint among the words related to the emotion being analyzed at the moment. This midpoint defines the exact location of the group in the space just created.

As the set of documents used in this example is too small (seven short texts), it was not possible to build the matrix of centroids. In Table 7 is shown part of the matrix generated in one the experiments carried out. On this table, because the original one contains 50 dimensions, it will be shown only the first six coordinates of each group.

Table 7. Identified Emotions for Each News

Alegria (joy)	*Desgosto (disgust)*	*Medo (fear)*	*Raiva (anger)*	*Surpresa (surprise)*	*Tristeza (sadness)*
0.0198	0.0054	0.0022	0.0011	0.0000	0.0035
0.0136	-0.0103	-0.0024	-0.0009	0.0000	-0.0052
0.0030	0.0088	-0.0014	-0.0005	0.0000	0.0009
0.0068	-0.0027	0.0000	-0.0003	0.0000	-0.0023
-0.0018	0.0011	0.0025	0.0040	0.0000	0.0114
-0.0031	0.0010	0.0004	-0.0028	0.0000	-0.0139

Finally, the last step is to use matrix V_T and the last matrix created, which contains the coordinates of the groups of emotions. Through cosine similarity [13], described on (1), we determine the distance between the document and the center of each group of emotions:

$$Sim(Gn, Dm) = \frac{\sum_i (W_{D_n} i * W_{G_m} \mathrm{j})}{\sqrt{\sum W^2_{D_n\, i}} * \sqrt{\sum W^2_{G_m \mathrm{j}}}} \qquad (1)$$

where:

Dn corresponds to the analyzed document *n*;
Gm corresponds to the group *m* being considered;
W_{Dni} e W_{Dnj} are the coordinates *i* and *j* for the document *n*, respectively;
W_{Gmi} e W_{Gmj} are the coordinates *i* and *j* for the group *m*, respectively.

The results obtained through (1) range from -1 to 1, where -1 or 1 mean the two points are exactly the same within the space, and 0 means that both points are totally distant from each other. The closer the points are, the more the document can be considered inside the group.

In order to facilitate the visual analysis of the results, they were multiplied by 100. Thus, our results vary between -100 and 100.

The Identification Process. The first step (training) creates a "space of words". The space is created using a large number of short texts (section 3.2 presents an experimentation with such a space). Then, this space may be used to identify emotions in a new short text.

In order to understand the identification process, assume this new text: “Novo terremoto volta a sacudir região central da Itália: Tremor ocorreu na região de Abruzzos. Área atingida fica próxima à cidade de L'Aquila”[2]. The first step preprocesses the text using the same procedure done in the training step. The second step produces a vector q to be attached to the term-document matrix (last column in Table 3), meaning the frequency of each term in the document.

The next step inserts the new text in the space produced using SVD. To do that, one should calculate the coordinates of the new text in the space. This is done by evaluating the equation (2):

$$V = q^T * U_k * S_k^{-1} \tag{2}$$

where:

q^T corresponds to q transposed;

U_k corresponds to the coordinates of each term inside the space (produced in the training process), where K is the number of dimensions (arbitrarily defined as 50);

S_k corresponds to the matrix S (calculated in training) and not used until now.

The vector V may be used now to calculate the cosine similarity of the new text with each of the six groups of emotions.

Table 8. Identifiyed Emotions in the New Text

Text	*Alegria (joy)*	*Desgosto (disgust)*	*Medo (fear)*	*Raiva (anger)*	*Surpresa (surprise)*	*Tristeza (sadness)*
news	-2.0569	20.7262	-2.4197	-3.2550	0.7031	17.0053

Results in Table 8 shows that *disgust* and *sadness* are the emotions with higher scores.

The approach was evaluated in several experiments, using different “types” of short texts (news and tweets are examples). The results presented in the next section were obtained using short news.

3.2 Experimental Results

The method presented in section 3.1 was implemented in java. A corpus with 1002 short news extracted from www.globo.com was created. A set of 700 news was used for training and 302 for testing. A group of 13 volunteers (researchers, PhD students and undergraduate students) analyzed the news, indicating the most important emotion presented in each text. Each participant evaluated around 25 short news and each short news was evaluated by 3 participants.

[2] Translation into English: “New earthquake shaking around central Italy: Tremor occurred in the region of Abruzzi. Affected area is near the town of L'Aquila.”

Table 9. Results for Each Emotion

Emotion	# of texts	# of occurrences correctly identified	Accuracy
Alegria (joy)	116	69	59%
Desgosto (disgust)	78	60	77%
Medo (fear)	20	16	80%
Raiva (anger)	18	9	50%
Surpresa (surprise)	7	6	86%
Tristeza (sadness)	63	45	71%

Table 9 shows the results for each emotion. It is important to highlight that the number of short texts for each emotion is not controlled a priori. The corpus of texts (1002) was collected in the same day, from different contexts (sports, politics, economy, etc.).

These first results show that some improvements are achievable. For instance, the list of words used to represent each emotion should pass for another process of verification, trying adding or removing words without relation with some emotion.

4 Conclusions and Future Work

The main contribution of this paper is a method to identify emotions in short texts for Brazilian Portuguese. Each text is processed using an algorithm based on the Latent Semantic Analysis (LSA) theory.

Many interesting applications could be improved by this research. For instance, the recognized emotions may be used to animate an avatar that plays the role of a television news program host.

We are developing different strategies to improve our results. First, we are studying the impact of regionalisms in the identification process. Texts written by people from the north have different emotions than texts written by people from the south?

We are also working to implement another algorithm in order to substitute de LSA-based algorithm used until now.

References

1. Fehr, B., Russel, A. J.: Concept of Emotion Viewed From a Prototype Perspective. Journal of Experimental Psychology, 464-486, Washington, (1984)

2. Gazzaniga, M. S., Heatherton, T. F.: Ciência Psicológica: Mente, Cérebro e Comportamento. Artmed. (In Portuguese), Porto Alegre, (2005)
3. Strongman, K. T.: The Psychology of Emotion. Chichester: John Wiley & Sons Ltd., (2003)
4. Ekman, P., Friesen, W. V.: Facial Action Coding System. Palo Alto: Consulting Psychologists Press, (1978)
5. Strapparava, C., Mihalcea, R.: Learning to Identify Emotions in Text. In: ACM Symposium on Applied Computing, pp. 1556-1560, (2008)
6. Pang, B., Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In: Proceedings of the 42nd ACL, pp. 271-278, (2004)
7. Cornell University (CU) Computer Science (CS). Sentiment Analysis. CS 40th anniversary Symposium, p. 26-27, (2005)
8. Landauer, T. K., Foltz, P. W., Laham, D.: Introduction to Latent Semantic Analysis. Discourse Processes 25, pp. 259-284, (2005)
9. Deerwester, S., Dumais, S., Landauer, T., Furnas, G., Beck, L.: Improving Information Retrieval with Latent Semantic Indexing. In: Proceedings of the 51st Annual Meeting of the American Society for Information Science, v.25, pp. 36-40, (1998)
10. Bradford, R.: Why LSI? Latent Semantic Indexing and Information Retrieval. Content Analysis. Agilex Technologies, Inc. Chantilly, Virginia, (2003)
11. WEKA: Data Mining Software in Java. The University of Waikato
12. Snowball Stemmer. http://snowball.tartarus.org/
13. Garcia, E.: Cosine Similarity and Term Weight Tutorial. http://www.miislita.com/ information-retrieval-tutorial/cosine-similarity-tutorial.html, (2006)
14. Grossman, D. A., Frieder, O.: Information retrieval: Algorithms and Heuristics. Second Edition. Springer, The Netherlands, (2004)