International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:12, No:10, 2018

Emotional Analysis for Text Search Queries on Internet

Gemma García López

Abstract**—**The goal of this study is to analyze if search queries carried out in search engines such as Google, can offer emotional information about the user that performs them. Knowing the emotional state in which the Internet user is located can be a key to achieve the maximum personalization of content and the detection of worrying behaviors. For this, two studies were carried out using tools with advanced natural language processing techniques. The first study determines if a query can be classified as positive, negative or neutral, while the second study extracts emotional content from words and applies the categorical and dimensional models for the representation of emotions. In addition, we use search queries in Spanish and English to establish similarities and differences between two languages. The results revealed that text search queries performed by users on the Internet can be classified emotionally. This allows us to better understand the emotional state of the user at the time of the search, which could involve adapting the technology and personalizing the responses to different emotional states.

Keywords**—**Emotion classification, text search queries, emotional analysis, sentiment analysis in text, natural language processing.

I. Introduction

EMOTIONS are indispensable in the development of human relationships and play a very important role in communicative processes. They have been the object of studies since the appearance of Darwin's book, *The Expression of emotions in Man and Animals* in 1873, to the application nowadays of sentimental analysis, "the computational study of opinions, sentiments and emotions expressed in text" [1]. Emotion detection in text allows us to distinguish between objective language (unbiased, not influenced by the writer's thoughts), and subjective language (influenced by the writer's opinions, emotions, or thoughts) [2], and is used especially in cognitive science and affective computing [3]. The detection of sentiment or emotion can be used in areas such as negative-positive polarity classification [4] and word emotional identification [5].

The recognition of emotions in search queries could be a great contribution to the use that society makes of the technology available. This research aims to approach the more personal side of a process we do every day: search on the Internet. If the search queries express emotions, the personal assistants, companies and search engines, could know more about the person who performs them. On the contrary that happens in emotional marketing, where we seek to awaken emotions, in this case we could know the emotions of the person and use that emotional information to improve human-computer interaction.

Through the use of search engines, users find information on the Internet. When conducting those searches, users experience emotions just as they do other everyday activities. Therefore, neither the information they search, nor the words they use to search, will be the same in a person who is feeling joy as someone who is sad or is feeling fearful. A user searching for "low calorie salmon recipes" is not feeling the same as another user looking for "diabetes symptoms".

Until now, search engines such as Google, Yahoo or Bing, and personal assistants such as Siri, can customize answers thanks to the information they have about their users. But also knowing the emotions that their users feel at the time of the search would be very valuable to offer more human and personalized content and answers that are based on these emotions.

II. Object and Method

With this exploratory study, we want to check if, in addition to the well-known intentionality of finding information, search queries offer emotional information through the words used. If so, emotions could be an important part of the search process, a key factor to include that could affect the use we make of the Internet. For it, Categorical Model and Dimensional Model [5] are applied to understand the emotional value of search queries. The main objective is to discover if search queries, despite the scarcity of words they use, can express emotions. And if so, by determining if the search queries express positivity or negativity, if they contain positive or negative emotions, we are at the same time approaching the emotions that the person performing the search manifests. To achieve this goal, the first step is to determine a positive/negative/neutral classification to help to define emotions more accurately. Second, for the text analysis and the understanding of the emotional content, categorical models will be applied through LIWC (Linguistic Inquiry and Word Count) and Tone Analyzer, and the dimensional model through ANEW (Affective Norms for English Words).

A. Study 1: Positive, Negative or Neutral Classification

To do this analysis, we will use search queries suggested by Google. The queries are made through keywords or phrases that a user uses to look for information. "[...] It is a word that is used as a reference point for finding other words of its kind or any information regarding those words" [6]. In this case, we will use Long Tail keywords, a keyword phrase with at least

G. G .L. Orcid (ID 0000-0002-8929-0563) is a PhD candidate in Complutense University of Madrid, Faculty of Information Sciences (Journalism department). Av. Complutense, S/N, 28040 Madrid, Spain (phone +34 913 94 22 20, e-mail: gglopez@ucm.es).

three words because they are more specific [7].

To select the search queries that we are going to use, we search some keywords on random subjects that are of interest on Google. Google search engine provides related results in the search bar and in the related searches section at the bottom of the page [8]. From the suggestions shown by Google, we selected those with more text for our analysis. The search queries follow a 5W1H model to extract information [9]. We also added some queries that do not follow these parameters. The same procedure is used for both English and Spanish (Annex I).

1. Development

Text Analytics from Microsoft Cognitive Services is used to classify the search queries obtained [10]. The Text Analytics API[2] is built with Azure Machine Learning, using advanced natural language processing techniques. This tool allows analysis of text and without needing knowledge of programming to use it. It can be used for research, marketing or personal purposes. Microsoft's Cognitive Services API has been used in other research projects [11], and it can be used to indicate positive-negative polarities in the sentiment analysis [12].

We have analyzed the results of 20 search queries on different topics. The searches had the same context, meaning and intention in both languages, changing only some typical expressions of each language. During the evaluation process, we introduce the search query in the tool, first in English and then in Spanish, for a simultaneous comparison of the results. Through the analysis of text, the tools directly show the percentage.

For the analysis of the results, the tool establishes 1% as very negative and 100% as very positive. A red bar represents percentages near 1%, while those close to 100% are shown in green, and the percentages close to 50% are represented in orange.

To be more precise in the analysis, we added to the classification a "neutral range" for those percentages that are between 45% and 55%. This choice is justified because the percentages close to the equator cannot be considered negative or positive at the same level as the others around 1% or 100%. Therefore, the percentages from 1% to 45% are considered negative, 45% to 55% are neutral, while 55% to 100% are positive.

Data extracted from Microsoft Cognitive Services Text Analytics show that of the 20 questions analyzed in Spanish, 15% (3 queries) are in the negativity ranges, 40% (8 queries) yield neutral data, and the 45% (9 queries) are considered as positive. Of the data analyzed for the English queries, 25% (5 queries) are negative, 20% (4 queries) are neutral and 55% (11 queries) are classified as positive (Fig. 1). See complete results in Annex II.

[2] The API returns a numeric score between 0 and 1. Scores close to 1 indicate positive sentiment and scores close to 0 indicate negative sentiment. Sentiment score is generated using classification techniques. The input features of the classifier include n-grams, features generated from part-of-speech tags, and word embedding. English, French, Spanish and Portuguese text are supported" (Microsoft Cognitive Text Analytics API).

2. Results

The majority of queries in both languages, English and Spanish, were classified as positive. However, the difference between the number of queries classified as positive and neutral is small. For example, in the consultations in Spanish, the difference between the number of queries classified as neutral and positive was 5%, while 40% were neutral and 45% positive. In the consultations analyzed in English the difference was broader, with 20% of the search queries analyzed within the neutral range and 55% in the positive range.

If we compare the results between both languages, only in search query 4 does the tool show disparate data. It offers a negative percentage for the query in Spanish, while it is positive in English. To make a double comparison, we performed the same query again but with a literal translation of the question in Spanish. The results continue to be positive. This difference can be due to the fact that in Spanish the preposition *sin* has a negative connotation meaning *absence of*. To check if this preposition has the same effect in English, we make the same query using *without*. Results in English remain positive. There is a clear discrepancy between the conclusions drawn in both languages. On the other hand, we found that 55% of the queries show the same positivity, negativity or neutrality data in both languages. In none of the cases can we speak of identical percentages for English and Spanish, but of percentages that are between the established ranges. If we analyze the queries that are between the positivity ranges in both languages, we find that those containing the adjective *mejor/the best* are always classified as positive (queries 1, 13 and 19). In addition, the tool shows the adjective highlights. Something similar happens with question 14. This query includes the adjective *buen*/*good* with positive meaning in both languages.

Query 7 is identified by the tool as positive. Although at first we could associate this sense to the verb *giveaway* and word *gift*, Text Analytics focuses on the word *novio*/*boyfriend*. The union between both concepts gives the query a positive character.

In query 5, the tool shows data of absolute negativity (0% in Spanish and 3% in English). This is due to the use of the word *dolor de cabeza* and *headache*, whose connotation is negative. On the opposite side, with absolute positivity, we find query 19, analyzed previously (100% in Spanish and 85% in English). Analyzing disparate questions with ranges of neutral - positive or neutral - negative, find some very interesting queries. In question 15, the tool shows 76% for the query in Spanish and 46% for English; this may be because the concept of 'cheap' in Spanish is very positively associated with saving money.

In query 9, the data are 36% for Spanish and 46% for English. This search query uses the word *flu* and *vacuna*/*shot*, both terms with negative connotation, especially in the case of the flu, as it is associated with illness. The tool in the analysis of the Spanish focuses its attention on words, *vacuna* and *gripe*. In the case of the query conducted in English, the tool focuses on the word *flu*. This may be the cause of not having

International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:12, No:10, 2018

classified both as negative.

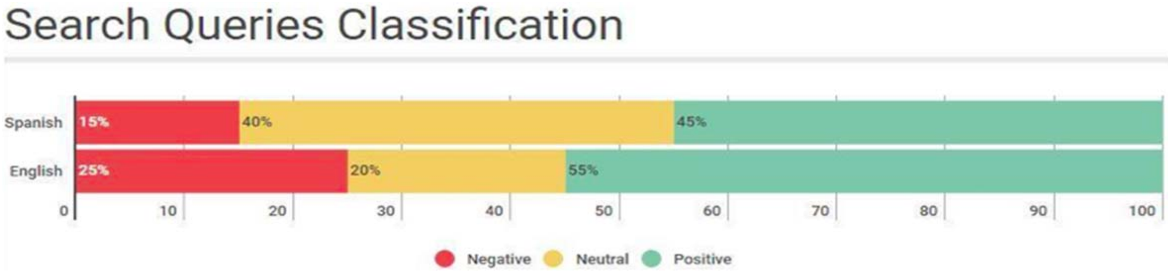

Fig. 1 Classification of results by percentages

As examples of queries classified as neutral in both languages, we have search queries 12 and 18. In the first, the most important word is *Tokyo,* it is an informational query; thus, that means that its classification is correct, and there is not subjective language. In the case of query 18, '*why trump won the elections?'* the intention of the message can be very subjective, but if you objectively analyze the words used, the classification is also correct.

There are other queries that create confusion and are difficult to interpret, for example query 10: *como eliminar mi facebook para siempre*/*eliminate facebook account permanently*. In this analysis, the tool classifies the query in Spanish as neutral and the query in English as negative. The conclusion in English could be justified because eliminate has a negative meaning in this sentence, however in key phrases establish that Facebook account is the most important point. In the case of Spanish, the verb *eliminar* and the adverb temporal (pron + adv) *para siempre*, should be sufficient indications for a negative qualification.

As an important point for the analysis, we must keep in mind that the information that we are going to provide to this tool is limited, compared to when we analyze text snippets, movie reviews, and other documents, where the word density is much higher. Text analysis tools are able to analyze phrases, text fragments and paragraphs; however, when we analyze search queries, we work with a few words, with an individual action occurring at a specific time in specific circumstances.

3. Discussion

The Microsoft Cognitive Services Text Analytics tool can establish a negative-positive ranking; however, researchers need to evaluate the differences based on linguistic and cultural parameters. Text Analytics is a tool that works with polarities and, although it uses orange in the central percentages, it does not work directly with a neutral range. This neutral area is interesting for analyzing results. In other text analysis tools such as Bitext, we have verified that this range of neutrality exists. Through the words, the machine is able to interpret the positive or negative connotation of some words. It is interesting to find that search queries, despite their low use of words, can also be classified as positive or negative. This first contact with the message also brings us closer to the user who transmits it. Through the words used, we can have a first contact with the transmitter and know if that message contains a positive or negative charge. This analysis is complex because if we look at its structure, in search queries we do not always find functional words, especially in Spanish, where the use of pronouns in queries is irrelevant. Thus, we find the information through the content words [13].

Search queries are a difficult element to analyze and classify. The first reason is the scarcity of text for the analysis, and the second is that the words used in the queries do not always have the intention that is first perceived. If we also add cultural differences, we will find other elements that have an affect such as the grammatical construction of a sentence or the expressions of each language. When the user performs a search, he tries to obtain a response by investing as little effort as possible and uses just words. With the approach to natural language thanks to personal assistants like Google Assistant, Siri or Cortana, our searches are more and more natural, but the information provided remains scarce. Detecting and differentiating the positive or negative sense in search queries could help search engines and personal assistants customize the content of the search results that are displayed to those questions. However, the big difference is not only know the positivity or negativity, but also the emotions that the user expresses in his searches.

B. Study 2: Emotional Classification

"Emotion detection is a NLP application that benefits from being able to distinguish subjective from objective language" [2]. Emotion detection in text is a difficult task that requires the application of comprehensive content analysis, natural language processing (NLP) and automatic learning techniques [14]. There are two models for representing emotions: categorical model and dimensional model [15]. A categorical model assumed that there are basic categories or emotions [16] such as anger, disgust, fear, joy, sadness and surprise. This model uses emotion-denotes words, or category labels to label each emotion with specific characteristics. The Dimensional Model defines emotional states as a space of two or three dimensions. Emotions have a place in that space. These types of models explore scales of classification for each dimension [17], and use tools such as Self Assessment Manikin (SAM)[3]

[3] a non-verbal pictorial assessment technique that directly measures the pleasure, arousal, and dominance associated with a person's Affective reaction to a wide variety of stimuli.

[18]. Currently, there are a large variety of dimensional models, each with its particular form of representation. Some examples are the circumplex model of affect [19], Plutchik's emotion wheel, a model based on the affective dimensions of kindness and excitation [20], or the three-dimensional PAD (Pleasure-Arousal-Dominance) representation [21].

1. Development

The goal of this study is to determine if search queries can provide specific emotional information through textual analysis [2]. To do this, we use the sample from the previous analysis with search queries in English and Spanish. We divided the search queries in English into positive and negative according to the classification extracted in study one, and performed the same procedure for the Spanish queries. Each query is given a number to identify them in a later analysis.

Categorical Model Application

Through a categorical representation, emotions can be labeled for classification [15]. Therefore, for the application of this model in our analysis, we used Linguistic Inquiry and Word Count (LIWC), a text analysis program that counts words and classifies them into psychological categories. Among its many categories, LIWC can identify emotion in language accurately [22]. LIWC has been previously used for the detection of emotions during text-based communication removing nonverbal cues [23], and is enabled for different languages analysis. To do our analysis, we created a .txt file for each of the search queries (SQ for English and CB for Spanish) and loaded the files into LIWC 2015.

2. Results

We found that the results for the emotional categories (Annex III and Annex IV), are very limited in both languages, English and Spanish. In the classification of queries in English, only some data are shown in positive queries but no data in negative ones. In the case of queries in Spanish, we can only extract a percentage in a query classified by LIWC as negative.

The data presented for the Spanish query show as negative, a search query established as positive in study 1. All the affective charge, (14,29 in CB10.txt) falls on the negative emotion. In the case of the queries in English, all those that have been classified into the category of positive emotion, coincide with the data extracted in study 1 where they were also classified as positive.

The data shown by the LIWC tool for the search queries in both languages is scarce and not very relevant, a fact that can be due to the shortage of text since the search queries do not exceed 10 words in any case. To counteract the scarcity of information, we used another tool for emotion analysis in text: Tone Analyzer of IBM Watson, a tool that "uses linguistic analysis to detect three types of tones from written text: emotions, social tendencies, and writing style" (IBM Watson Developer Cloud).

Tone Analyzer assesses three main categories: emotional, social, and language, and sets a score from 0 to 1: not likely present <0.5, likely present 0.5 - 0.75 and very likely present >0.75. This tool offers data about five specific emotions: anger, disgust, fear, joy and sadness. Despite being a powerful tool, it still does not offer different languages services. We made a test exclusively for the search queries in English with the aim of making a comparison using the data of study 1.

English Queries	Anger	Disgust	Fear	Joy	Sadness
SQ 1	0.06	0.02	0.07	**0.82**	0.09
SQ 2	0.03	0.12	0.17	**0.41**	0.36
SQ 3	0.06	**0.35**	0.10	0.23	0.34
SQ 4	0.02	0.09	0.04	**0.74**	0.14
SQ 5	0.01	0.08	0.05	**0.76**	0.16
SQ 6	0.21	**0.38**	0.10	0.28	0.12
SQ 7	0.03	0.07	0.10	**0.57**	0.28
SQ 8	0.14	0.07	0.10	**0.66**	0.10
SQ 9	0.04	0.03	0.13	0.38	**0.51**
SQ 10	0.13	0.05	**0.54**	0.14	0.21
SQ 11	0.06	0.03	0.19	**0.54**	0.23
SQ 12	**0.44**	0.03	0.10	0.08	0.40
SQ 13	0.21	**0.29**	**0.29***	0.20	0.10
SQ 14	0.39	0.05	**0.52**	0.01	0.17
SQ 15	0.18	0.04	0.05	0.05	**0.75**
SQ 16	0.09	0.06	0.27	**0.35**	0.31

Fig. 2 Tone Analyzer results for English Search Queries

Comparing the results shown by Tone Analyzer with the results of study 1, we found some very interesting data. Of the 11 search queries that were classified as positive in study 1, Tone Analyzer ranks the majority (7 queries) as Joy emotion. While, the other two queries are classified as disgust, one as fear and one as sadness. The search queries "can you eat tofu straight from the package" and "how to make a personal budget on excel" were classified as disgust but with a range <0.5, so the data are not representative.

The query *when is the second season of stranger things coming out* was ranked in its classification as fear, this may be due to the presence of *stranger things* and shows how the tool has analyzed a set of words but not the intention of the sentence as a whole. The query *who won best director oscars 2015* has been classified as sadness, something that contradicts the results of study 1 where the query was classified as positive with 85%.

For search queries classified as negative, only one of them showed contradictory data but with a very low rank. *Restaurants near me open now* was rated negative in study 1 but Tone Analyzer placed it in joy emotion with range not likely present <0.5, so the data are not representative. Meanwhile, *how do driverless cars work* was rated negative in study 1 and Tone Analyzer placed it with a score of 0.44 in anger emotion. It is the only query classified like anger. Comparing the data extracted by LIWC and Tone Analyzer, we could observe that of the six search queries classified as positive by LIWC, only four appear associated with the joy emotion in Tone Analyzer. However, in query 6 and query 9, the results are contradictory. The results shown by the LIWC tool for the application of the categorical model were insufficient for this object of study. The results shown by Tone Analyzer were more precise and we could find many similarities with study 1. However, the limitation of text and the scarcity of resources for Spanish are two handicaps for this type of analysis.

Dimensional Model Application

To test the dimensional model and apply it to our study, we used the Affective Norms for English Words (ANEW) [24]. ANEW is “a set of normative emotional ratings for a collection of English words (N=1,035), where after reading the words, subjects report their emotions in a three dimensional representation. This collection provides the rated values for valence, arousal, and dominance for each word rated using the Self Assessment Manikin (SAM)” [15]. For Spanish Adaptation of ANEW, the evaluations were done in the dimensions of valence, arousal and dominance using (SAM), the Self-Assessment Manikin [25].

ANEW is the result of psycholinguistic experiments in individual and mixed groups of women and men. It uses three dimensions of affection for the classification of the words: Valence, Arousal and Dominance, with a punctuation of three numbers in the range 1-10. Negative emotions such anger and fear can be associated with negative valence, and joy with positive valence. Arousal measures the intensity of an emotion. For example, rage may show higher excitement than anger, although both are negative emotions. And finally, dominance shows the dominant nature of an emotion.

English Queries	Words in ANEW	Valence Mean (SD)	Arousal Mean (SD)	Dominance Mean (SD)
SQ1	Not found	X	X	X
SQ2	world	6.50 (2.03)	5.32 (2.39)	5.26 (2.47)
SQ3	eat	7.47 (1.73)	5.69 (2.51)	5.60 (2.12)
SQ4	gift idea	7.77 (2.24) 7.00 (1.34)	6.14 (2.76) 5.86 (1.81)	5.52 (2.54) 6.26 (2.00)
SQ5	gift idea	7.77 (2.24) 7.00 (1.34)	6.14 (2.76) 5.86 (1.81)	5.52 (2.54) 6.26 (2.00)
SQ6	Not found	X	X	X
SQ7	world	6.50 (2.03)	5.32 (2.39)	5.26 (2.47)
SQ8	good leader	7.47 (1.45) 7.63 (1.59)	5.43 (2.85) 6.27 (2.18)	6.41 (2.05) 7.88 (1.60)
SQ9	Not found	X	X	X
SQ10	Not found	X	X	X
SQ 11	Not found	X	X	X
SQ12	car	7.73 (1.63)	6.24 (2.04)	6.98 (2.06)
SQ13	headache medicine	2.02 (1.06) 5.67 (2.06)	5.07 (2.74) 4.40 (2.36)	3.60 (1.98) 4.70 (1.91)
SQ14	finger	5.29 (1.42)	3.78 (2.42)	5.05 (1.70)
SQ15	Not found	X	X	X
SQ16	restaurant	6.76 (1.85)	5.41 (2.55)	5.73 (1.41)

Fig. 3 Words extracted from search queries in English using ANEW

This model has been used in other studies applied to social media [26] and microblogs. There are also examples of the application of the Spanish Adaptation of the Affective Norms for English Word, specifically in the study of hotel opinions [27]. To do our analysis, we have extracted the words from each of the search queries in both languages and we looked for their classification in ANEW, using the data from valence, arousal, and dominance, with scores for averaged across all subjects.

Spanish Queries	Words in the Spanish adaptation of ANEW	Valence Mean (SD)	Arousal Mean (SD)	Dominance Mean (SD)
SQ1	Not found	X	X	X
SQ2	Not found	X	X	X
SQ3	novio	7.49 (2.14)	6.70 (2.41)	5.13 (2.17)
SQ4	Not found	X	X	X
SQ5	mundo	6.57 (1.96)	6.12 (2.10)	3.90 (2.28)
SQ6	Not found	X	X	X
SQ7	Not found	X	X	X
SQ8	Not found	X	X	X
SQ9	Not found	X	X	X
SQ10	Not found	X	X	X
SQ11	Not found	X	X	X
SQ12	Not found	X	X	X

Fig. 4 Words extracted from search queries in Spanish using ANEW

3. Results

After analyzing the words extracted from the English search queries in ANEW, we found that the words of the positive queries had a high valence degree, higher than 5 points and also arousal higher than 5 points. This data corroborated that these words are associated with positive emotions as we discovered in study 1. However, in the case of negative search queries, in the cases that have been analyzed, they did not show a low valence as expected. Only the word *headache*, classified in negative queries showed negative valence with arousal above 5 points. Dominance ranges did not provide relevant information for this object of study. The word with the highest degree of valence was *gift*, present in two search queries classified as positive with a percentage higher than 80%. *Leader* was the word with greater arousal and dominance and was present in a search query classified as positive with 89%. However, in the same query, the word *good* was classified as positive, but analyzed individually it showed a moderate arousal (5.43), a disconcerting fact about an adjective. In the case of words in Spanish, the number of words analyzed is very low. The adaptation of ANEW to Spanish does not contain all the words present in the original ANEW, and of the words used in the search queries, most are not present. Of the two analyzed words, we could extract that they present a positive valence and both words were extracted from queries of positive searches in study 1.

In both Figs. 3 and 4 was present the word *world* - *mundo* with a similar valence. The word *world* has a valence of (6.50) and *mundo* (6.57). In the case of arousal, the score is higher in the case of *mundo* and for dominance, *mundo* shows a score of (3.90), compared to (5.26) *world*. This shows that a word in both languages can have the same valence, be in a positive emotional range but show different intensity. The context and culture could determine the difference in these ranges of arousal and dominance.

This dimensional model can serve as a complement to emotional analysis in text and determine the degree of emotion of words, but alone cannot show what kind of emotion it expresses. To obtain representative data in our analysis, the words analyzed separately do not have data of conclusions, but do support other models of analysis.

ANEW is a resource that does not contain all the words in the dictionary, and in the case of its adaptation to Spanish, the number of words represented is still scarce. The problem we detect is the scarcity of resources and tools for the analysis of emotions in Spanish text, and therefore, for comparison with other languages. Its differential value is that the ranges of valence, arousal and dominance can be used to analyze contextual, linguistic and cultural differences in different languages.

Each model provided emotional information according to its structure. Comparing the results of those search queries present in the different representations, we found for example that the query 2, was classified as positive in study 1, contains the word *world* with a valence of 6.50 according to ANEW and LIWC also classified it as positive.

Something similar happened with query 4 and query 5, classified as positive in study 1, as joy by Tone Analyzer and it contain two words with a valence greater than 7.00 according to ANEW. Query 7 was also classified as positive and was included in the joy emotion. According to ANEW, the word *world* has a valence of 6.50. LIWC also classified it as a positive emotion. This word was also present in the Spanish results, with a score of 6.57 and classified as positive in study 1.

For query 8 and query 9, established as positive in study 1, LIWC shows positive emotion but Tone Analyzer shows joy in query 8 and sadness in query 9. For query 13, Tone Analyzer establishes a rating of fear and ANEW shows a valence less than 5, which is related to a negative emotion. This query was also classified as negative in study 1. We should remember that Tone Analyzer established this query as fear but obtained the same score, 0.29, also for emotion disgust.

In general, of the 11 positive search queries in English extracted from study 1, six have been corroborated by some of the models and classified with the joy emotion. Of the five negative queries, four have been classified under negative emotions, supporting the first classification.

III. Conclusion

The findings suggest that search queries can be classified emotionally. Using different tools and dictionaries available, we can know if search queries are positive, negative or neutral, and if they show emotions such as joy or sadness. Words that contain search queries can be analyzed to show an emotional state embodied in text.

However, using a single tool for detecting emotions in such a small percentage of text may not be enough. It requires applying different models to draw conclusions. Another handicap is the language limitation. Most tools only work in English, and in those that admit Spanish such as LIWC, the shortage of text of the consultations makes analysis difficult. In recent years, much progress has been made in analyzing emotions and their application in different disciplines, but there is still a long way to go in the case of text detection. It would be interesting to be able to analyze a search query as an independent unit as well as to perform analysis at the sentence level within a document.

Another important factor is the cultural and linguistic differences between languages when we analyze the emotional meaning of a search query. These differences cannot be captured in analysis tools yet. There are words specific to each language and others with double meaning that depending on the accompanying words, may have one meaning or another. The connotation of words in each language is essential to understand the meaning of the sentence and also the implicit emotions. Even within the same language, the interpretation of a word may be different. For example, in the UK, the word *tube* refers to the subway, while in America, it is assumed to be a TV. So if we want to know emotions, we have to interpret them within a context and culture.

Understanding the emotions of a user through text helps us to get closer to the human side of HCI (Human-computer interaction), and to analyze the sender and his message. This study could contribute to the understanding of the behavior of the search queries in different languages and the technology adaptation for different emotional states. Knowing the positivity or negativity of the search queries, and the emotions that the user manifests through the words they use, can be a fundamental piece in the algorithms of big companies like Google for the personalization of the answers.

Including the emotional factor in search queries is an advance for society and for technology. Search engines such as Google, Yahoo or Bing could use emotional information to offer emotionally personalized responses. Currently, if we look for information about premiere films, search engines show us the titles and the closest cinemas to see them. But, what if the search engine shows us first comedy movies if we are sad? Another example is queries that can be classified as fear, as is the case of *what to do if there is an earthquake*, which should be answered only with rigorous, not alarming information.

The detection of emotions could suppose an unprecedented filter for the content of any field. Using emotions in search processes would involve knowing the emotional state of the user to be able to offer always beneficial (but optional) answers. This could be of great help, for example in the field of health for patients with hypochondria, for detecting sadness or fear in their search queries and regulating the information that is accessed in the first place.

References

[1] Bing Liu (2010). Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing, Second Edition.
[2] Mulcrone, K. (2012). Detecting Emotion in Text. University of Minnesota–Morris CS Senior Seminar Paper.
[3] Picard. R. W. (1995). Affective Computing. MIT Media Laboratory; Perceptual Computing; 20 Ames St., Cambridge, MA 02139. M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 321.
[4] Bo Pang and Lillian Lee. (2008). Opinion mining and sentiment

analysis. Foundations and Trends in Information Retrieval, 2(1/2):1–135.
[5] Mac Kim, S. (2011) Recognizing Emotions and Sentiments in Text. University of Sydney.
[6] SemRush Company. (2013) What are Keywords? Simple Keyword Definition.
[7] Kusinitz S. (2014). The Definition of a Long-Tail Keyword (In Under 100 Words).
[8] Google Support (2017). Find related searches. Web.
[9] Buttry, Steve. (2011). The 5 W's (and How) of writing for the web.
[10] Frey, L., Botan, C., & Kreps, G. (1999). Investigating communication: An introduction to research methods. Chapter 9. (2nd ed.) Boston: Allyn & Bacon.
[11] Matsuyama, Bhardwaj, Zhao, Romero, Akoju and Cassell. (2016). Socially-Aware Animated Intelligent Personal Assistant Agent. ArticuLab, Carnegie Mellon University, Pittsburgh.
[12] Tetsuya Nasukawa and Jeonghee Yi. (2003). Sentiment Analysis: Capturing Favorability Using Natural Language Processing. IBM Research, Tokyo Research Laboratory. IBM Research, Almaden Research Center.
[13] Pennebaker, J. (2011). The secret life of pronouns. What our words say about us. Bloomsbury Press. New York.
[14] Shivhare, S. N., & Khethawat, S. (2012). Emotion detection from text. arXiv preprint arXiv:1205.4944. Department of CSE and IT, Maulana Azad National Institute of Technology, Bhopal, Madhya Pradesh, India.
[15] Calvo, R. A. & Kim, M. (2013). Emotions in text: dimensional and categorical models. Volume 29, Issue 3, pages 527–543. School of Electrical and Information Engineering The University of Sydney.
[16] Ekman, P. (1992). An argument for basic emotions. Cognition Emotion, 6(3), 169-200. University of California.
[17] Mac Kim, S. (2011) Recognizing Emotions and Sentiments in Text. University of Sydney.
[18] Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. Journal of behavior therapy and experimental psychiatry, 25(1), 49-59. University of Florida.
[19] Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and psychopathology, 17(03), 715-734. Columbia College of Physicians & Surgeons, New York State & Psychiatric Institute, Boston College.
[20] Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. Journal of experimental psychology: Learning, Memory, and Cognition, 18(2), 379-390.
[21] Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual. Current Psychology, 15(4), 505-525. University of California, Los Angeles.
[22] Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology, 29(1), 24-54.
[23] Hancock, J. T., Landrigan, C., & Silver, C. (2007, April). Expressing emotion in text-based communication. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 929-932). ACM. San Jose, CA, USA.
[24] Bradley, M. & P. Lang. (1999). Affective Norms for English Words (ANEW): Stimuli, Instruction Manual and Affective Ratings. Informe Técnico C-1, Gainesville, FL, The Center for Research.
[25] Redondo, J., I. Fraga, I. Padrón, y M. Comesaña. (2007) The Spanish adaptation of ANEW (Affective Norms for English Words) Behavior Research Methods, 39(3):600–605, Agosto.
[26] Arup Nielsen, F. (2011). (ANEW): Evaluation of a word list for sentiment analysis in microblogs. DTU Informatics, Technical University of Denmark, Lyngby, Denmark.
[27] Miranda, C. H., Guzmán, J., & Salcedo, D. (2016). Minería de Opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles. Procesamiento del Lenguaje Natural, (56), 25-32.