October 7-10 • Ceará • Brazil

34th Brazilian Symposium on DATABASES

SBBD|2019

PROCEEDINGS COMPANION

October 7-10 · Ceará · Brazil

34th Brazilian Symposium on DATABASES

PROCEEDINGS COMPANION

Steering Committee Chair
Bernadete Farias Lóscio (UFPE, Brazil)

Local Chair
José Maria da Silva Monteiro Filho (UFC, Brazil)

Program Committee Chairs
Full Paper: Carina F. Dorneles (UFSC, Brazil)
Short Paper: Fábio Porto (LNCC, Brazil)
Demos and Applications Chair: Robson L. F. Cordeiro (ICMC-USP, Brazil)
Thesis and Dissertation Workshop Chair: Jonice Oliveira (UFRJ, Brazil)
Tutorials Chair: Altigran Soares da Silva (UFAM, Brazil)
Short course Chair: Maria Cláudia Cavalcanti (IME, Brazil)
Workshop Chair: José Antônio Macedo (UFC, Brazil)
Thesis and Dissertation Contest Chair: Caetano Traina Jr. (USP, Brazil)
Graduation Student Workshop Chair: Ticiana Linhares (UFC, Brazil)

34th Brazilian Symposium on Databases

October 7-10, 2019
Fortaleza - CE - Brazil

Dataset Show Case

PROCEEDINGS

Promotion

Sociedade Brasileira de Computação – SBC
Comissão Especial de Banco de Dados (CEBD) da SBC

Organization

Departamento de Computação, Universidade Federal de Ceará– UFC

Program Chair

Mirella M. Moro

Table of Contents (Dataset Show Case)

Beyond Tears and Smiles with *ReactSet*:
Records of Users' Emotions in Facebook Posts

Mirela T. Cazzolato[1], **Felipe T. Giuntini**[1], **Larissa P. Ruiz**[2],
Luziane de F. Kirchner[3], **Denise A. Passarelli**[2], **Maria de Jesus Dutra dos Reis**[2],
Caetano Traina-Jr.[1], **Jó Ueyama**[1], **Agma J. M. Traina**[1]

[1] Institute of Mathematics and Computer Sciences
University of São Paulo - São Carlos, SP, Brazil,
[2] Department of Psychology
Federal University of São Carlos - São Carlos, SP, Brazil,
[3] Dom Bosco Catholic University, Campo Grande, MS, Brazil

`{mirelac, felipegiuntini}@usp.br,`
`{caetano, joueyama, agma}@icmc.usp.br`

Abstract. *Emotion and feelings recognition have been studied in a wide research range in the past decades. The advancement and spread of social networks and online applications, such as Facebook, Instagram, and Whatsapp has motivated the sharing of news and communications among users. Specifically, in social networks, users can give reactions related to different contents, which can provide meaningful clues to study emotions. In this work, we present ReactSet, a dataset composed of records of users' emotions for different news, along with some de-identified personal information and the time they took to analyze each news. All the news were selected by specialists from the areas of Psychology and Human-Behavior analysis. We describe ReactSet contents, present statistics, potential applications, and challenges regarding the collected data. ReactSet is publicly available for research use, under the Creative Commons license.*

1. Introduction

Social relations and networking are fundamental components of human life, historically linked according to limitations from time and space. However, the technological evolution, the Internet, and its diffusion of use have partially removed such restrictions. The emergence of Web technologies allowed services in virtual social networks as they exist in a non-virtual way. This notion of social networking and its methods of analysis has attracted significant interest and curiosity from the community at large in the last decades, especially from the areas of social and behavioral sciences. The interest derives from social networks providing a powerful abstraction of the structure and dynamics of different types of people or person-technology interaction. Social network analysis can indicate the study of virtual social structures and its effects to analyze social and cultural aspects.

People have increasingly prioritized agile, fast, and efficient communication. The form of peer communication in social networks and online applications such as Facebook, Instagram, Whatsapp, and Twitter has also been transformed and reduced each day to use fewer characters. In this virtual context, new notations formed by combining a set of alphanumeric characters and punctuation, such as :) , (+_+) , O_o , :(, indicate objects, gestures and especially expressions of opinions and emotions. These annotations
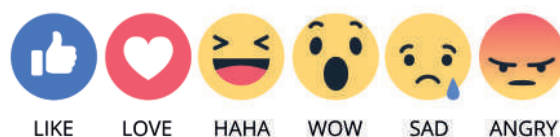
Figure 1. Six Facebook reactions, currently available at Facebook.

have been evolving and enhanced with graphical features such as colors (emojis) and animations (emoticons). Emoticons are small images or combinations of diacritic symbols that participants of social media can use to express feelings like excitement, anxiety, and anger. The content of posts, messages, and comments can include emoticons. Recently, the social network Facebook selected and restyled a set of more significant emoticons, and named them reactions. The reactions make it possible for users to react and express themselves emotionally to the published postings. Figure 1 shows the Facebook reactions Like, Love, Haha, Wow, Sad, and Angry.

Emotional facial expressions are powerful stimuli in the analysis of social interactions. Individuals use those expressions to identify common hazards, avoid conflicts, keep track of partners' emotional reactions, adjust their behavior depending on others' attitudes, learn to describe emotions verbally, and so on [Ekman et al. 1969, Ekman and Friesen 1971, Lerner and Keltner 2001, Russell et al. 2003, Spoor and Kelly 2004, Waller et al. 2008]. Although some theorists argue that such expressions are distinctly shaped at each culture [Fridlund 1994], others have pointed to their universal recognition [Ekman 1993, Frijda et al. 2000]. Charles Darwin was the first scientist to systematically evaluate the universality of facial expressions to recognize emotions, when in 1872 he wrote "The Expression of the Emotions in Man and Animals". Darwin's methods of study were based on the observation of the facial expressions of young children, as they interact with their peers, blind or sighted children, and people with mental disorders. From some experiments, Darwin found evidence showing that facial expressions correspond to specific emotions, regardless of the learning environment of each individual. In 1960, Paul Ekman continued the studies of this subject, which led him to conclude that there are a limited number of facial expressions that signaled emotions. In 1982 he postulated about the universality of expression of six basic emotions (anger, disgust, fear, joy, sadness, and surprise), and in 1990 he argued about the existence of eleven other universal emotions (fun, contempt, contentment, embarrassment, excitement, guilt, pride in conquest, relief, satisfaction, sensory pleasure, and shame). However, the latter still presents concerns regarding its universality [Wolf 2015].

The research on emotional expressions can be applied to different kinds of communication, and some of them are usually the most significant social media interactions. People are continually using social networks to express what they feel [Jibril and Abdullah 2013]. Facebook has about 2,38 billion monthly active participants, who spend most of their day online, making the virtual environment a useful source of evidence about what participants consider and feel about [Statista 2019, Vashisht and Thakur 2014]. Emoticons are currently the most broadly used media in the virtual environment [Oleszkiewicz et al. 2017], addressed by several scientific literature [Wegrzyn-Wolska et al. 2016, Wang and Castanon 2015, Oleszkiewicz et al. 2017, Huang et al. 2008]. Instead of describing emotions through words, emoticons can help to

provide hints for communication that may be in the virtual technology world, i.e., non-verbal clues [Vashisht and Thakur 2014]. The importance of this declaration has given rise to the concern of scientists, regarding how emoticons can be used in several fields, such as mental health issues, reactions to stressful occurrences, preferences or decisions, and multiple opinion polls [Luor et al. 2010, Gaspar et al. 2016].

Previous use of the data. Recently, several works on the data science field have explored how well emoticons or reactions can represent expressions of corresponding emotions in real environments. In this context, emoticons and reactions are non-verbal forms of expression in online social networks. In the work [Giuntini et al. 2019], the authors employed a questionnaire on Facebook, aimed at collecting different information of users regarding specific posts. A panel of specialists in Psychology and researchers in human behavior analysis selected a set of Facebook news, with a high number of reactions. These Facebook news were made available in a survey for Facebook users. Users responded to what reaction they attributed to the posting, the polarity of feeling (negative, neutral, or positive) and could choose up to two of Ekman's most significant basic emotions (anger, disgust, fear, joy, sadness, and surprise) [Ekman et al. 1969].

Giuntini *et al.* used part of the collected users' responses to news to address how people reacted to various relevant news, at a given moment, using Facebook reactions. In this work, we present an extended and complete version of the data reported in [Giuntini et al. 2019], called *ReactSet*. *ReactSet* is a new dataset with emotions and reactions of Facebook users regarding 34 news. It provides support for different analysis involving emotions. Moreover, the dataset can serve for data mining researches that seek to solve problems related to the study of feelings, temporal patterns analysis, subjectivity, and the treatment of missing and incomplete data.

Paper outline. The remainder of this paper is organized as follows. Section 2 describes the *ReactSet* dataset. Section 3 discusses the main applications and challenges we envision to use *ReactSet*. Section 4 details how *ReactSet* is publicly available for download, along with the description of its data files in the public repository GitHub[1]. Finally, Section 5 gives the conclusion of this work.

2. *ReactSet*: Reactions of Facebook Users

In this section, we describe *ReactSet*, a dataset of Facebook users' reactions to different news.

2.1. Data Collection and Preprocessing

ReactSet was collected by the application of a questionnaire, which was publicly open to answers on Facebook during the month of July 2017. In total, 409 participants answered the questions, 149 of them with complete responses, and 260 of them with partial answers. The employed questionnaire has two parts. The first consists of personal information, such as age, sex, and scholarity. The second part asks the user to analyze different news. A total of 36 news were presented to the users, each with an image and a description. For each news, users were asked to answer the following questions:

- *Question 1.* Which reaction would you give to this post?

---

[1]*ReactSet* is publicly available for download at *https://github.com/mtcazzolato/reactset*.

– *Answers*: Like, Love, Haha, Wow, Sad, Angry, using icons from Figure 1.
- *Question 2.* How do you classify this post?
    – *Answers*: Positive, Neutral, Negative.
- *Question 3.* Which is the most predominant emotion is this post?
    – *Answers*: sadness, joy, do not recognize, fear, disgust, anger, surprise.
- *Question 4.* Which is the second most predominant emotion in this post?
    – *Answers*: sadness, joy, do not recognize, fear, disgust, anger, surprise.

A group of 7 judges selected the news we used for the analysis. This group was composed of Master, Ph.D. students and specialists on human-behavior analysis. They first classified each news as containing one of the emotions: anger, disgust, fear, joy, sadness, and surprise. From the 36 news, judges selected the 24 most representative news (4 for each emotion), which presented the following characteristics [Giuntini et al. 2019]: (i) the majority of the judges found the same emotion in a single news; (ii) the majority of the participants found an emotion in the same news item. We summarize the judges' concordance for the 24 selected news in Table 1, along with the participants' concordance, separated by the observed emotion.

2.2. Data Description

Figure 2 presents the dataset schema, and Table 2 presents the tables and columns of *ReactSet*, along with their corresponding description. Three tables organize all available information:

- *PersonalInfo.* The users' personal information, and the dates in which the user answered the questions;
- *News.* The information regarding each news evaluated;
- *AnswerNews.* The answers of users regarding each news present in table News.
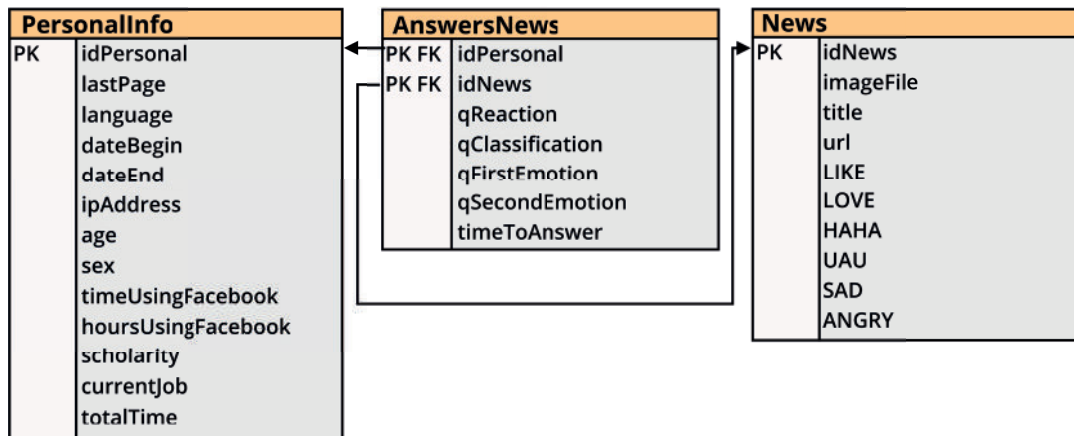
Figure 2. Schema of the *ReactSet* dataset.

In table *News*, we make available the number of reactions given by Facebook users for each available news, gathered in July, 2019. Notice that these reactions are those obtained by the Facebook post, given by the general public. The information is provided for 21 of 36 news. Figure 3 (a, b and c) shows three examples of news. Figure 3 (d) shows the collected Facebook reactions of the three news. We observe that the first news, a potential

Emotion	News ID	Participants' concordance (%)	Judges' concordance (%)
Joy	4	55.10	71.42
	10	97.96	85.71
	21	95.24	71.42
	24	90.48	85.71
Sadness	12	74.83	100
	13	85.71	85.70
	26	84.35	85.71
	27	78.23	71.42
Anger	2	63.27	71.42
	7	80.95	57.14
	23	44.22	57.14
	28	59.89	85.71
Disgust	16	23.80	28.57
	20	57.82	42.85
	30	78.91	28.58
	35	14.29	42.85
Fear	15	17.00	71.42
	29	14.97	71.42
	31	20.41	28.58
	32	7.48	57.14
Surprise	5	51.02	100
	11	70.09	57.14
	33	73.47	71.42
	25	69.39	85.71

Table 1. Users' and Judges' concordance of 24 selected news, 4 of each emotion.

terrorist attack, has received more negative reactions than the two other news, regarding sports and entertainment matters. Accordingly, we observed an agreement regarding the news content and the observed reactions.

Figure 4 (a) depicts the sex per age of the participants. From a total of 409 participants, 239 identified themselves as being a woman, 124 identified themselves as a man, 4 preferred not to inform their sex, and 42 of them did not provide an answer. We observe that women have higher age values, in comparison to men and other answers. Also, Figure 4 (b) shows that the majority of the participants are undergraduate students, followed by high school and master students. Figure 4 (c) shows the sum of reactions given by the participants, for each of the analyzed news. As we observe, the number of missing responses to the questions increased with the number of news presented by the participants. The reaction, polarity classification, and emotions of each participant according to the post can be studied in a range of applications, as we discuss next.

Table 2. Tables and columns of *ReactSet*.

Table	Column	Type	Description
PersonalInfo	idPersonal (PK)	int	Identifier of the person / answer
	lastPage	int	Last page of the questionnaire visited by the person
	language	string	Language of the questionnaire
	dateBegin	datetime	Date of when the person started answering the questionnaire
	dateEnd	datetime	Date of when the person ended answering the questionnaire
	ipAddress	string	IP address of the computer / device used to answer the questionnaire
	age	int	Age of the person
	sex	string	Sex of the person
	timeUsingFacebook	string	How long the person uses Facebook
	hoursUsingFacebook	string	Number of hours the person uses Facebook (daily average)
	scholarity	string	Complete education level
	currentJob	string	Current job situation
	totalTime	float	Total time the person spent to answer all questions
News	idNews (PK)	int	Identifier of the news
	imageFile	string	Image file name
	title	string	Title of the news depicted in the image
	url	string	URL of the news depicted in the image
	like	string	Like reactions of the news on Facebook
	love	string	Love reactions of the news on Facebook
	haha	string	Haha reactions of the news on Facebook
	uau	string	Uau reactions of the news on Facebook
	sad	string	Sad reactions of the news on Facebook
	angry	string	Angry reactions of the news on Facebook
AnswersNews	idPersonal (PK)(FK)	int	Identifier of the person / answer
	idNews (PK)(FK)	int	Identifier of the news
	qReaction	string	Reaction the person assigned to the news
	qClassification	string	Polarity classification the person assigned to the news
	qFirstEmotion	string	The most predominant emotion the person identified in the news
	qSecondEmotion	string	The second most predominant emotion the person identified in the news
	timeToAnswer	float	Time the person spent to answer the questions related to the news

(a) News ID 06.   (b) News ID 25.   (c) News ID 32.

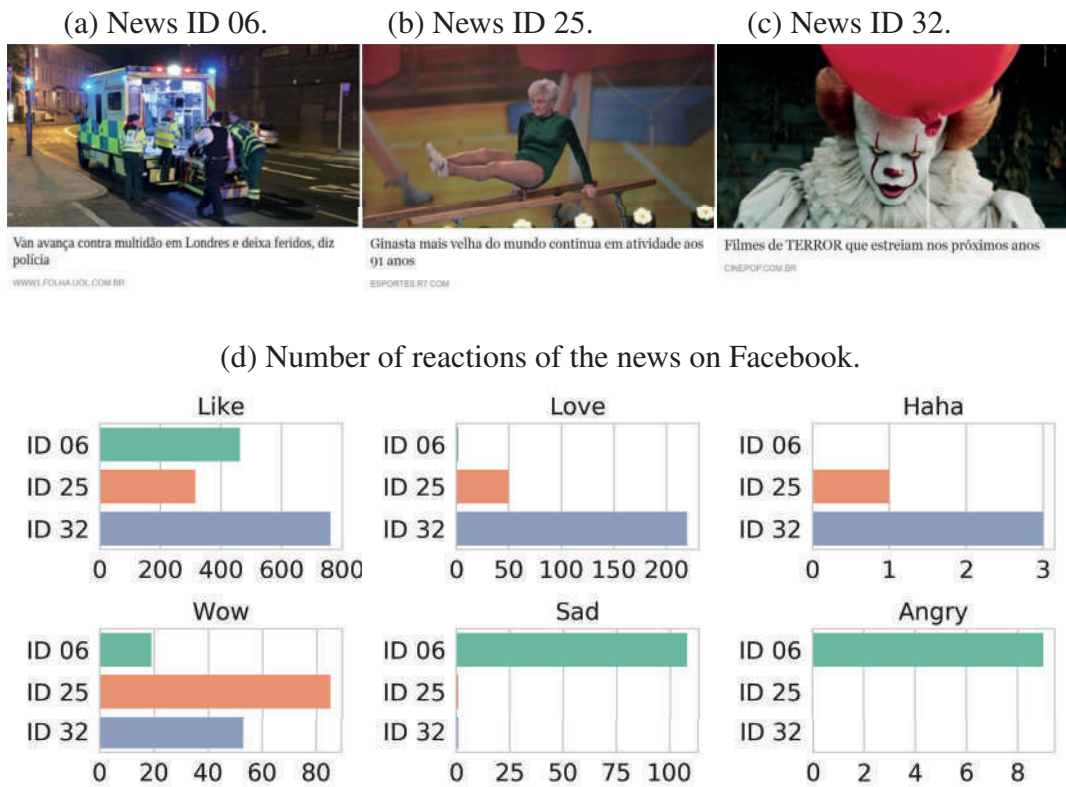(d) Number of reactions of the news on Facebook.

Figure 3. Example of three news present in *ReactSet*, and the number of reactions given by Facebook users.

## 3. Applications and Challenges regarding *ReactSet*

ReactSet opens research opportunities in various application contexts in the field of data science and analysis of feelings. We discuss potential opportunities following.

i. Temporality: For each question group, which corresponds to the questions regarding single news, *ReactSet* contains the information of how long the user took to assign the corresponding reaction, polarity classification, and emotions. This time information can allow the analysis of aspects such as excitement, motivation, and spontaneity of the participants when answering the specific news. For example, check if the user acted quickly or had to reflect on the image and title of the news.

ii. Missing data: Several researches have tried to develop new approaches so that algorithms of mining and machine learning can operate with good performance in databases with missing data. From 409 responses obtained in the survey, only 149 of the participants completed the survey. This opens up challenges, such as: (i) predicting the values of empty cells based on the user's previous responses; (ii) develop new algorithms or adaptation of existing ones to better classify feelings and emotions; (iii) address newly information, such as image content or textual news information to predict the reactions.

iii. Predictions: Considering that reactions have been used both to show affection, as well as opinion on brands and products, the prediction of reactions in social networks can help the development of more attractive, audience-guided, and positive impact ads.

iv. Small Data: For some problems, including emotions analysis, gaining knowledge

(a) Sex per age of the participants.   (b) Participants per scholarity level.
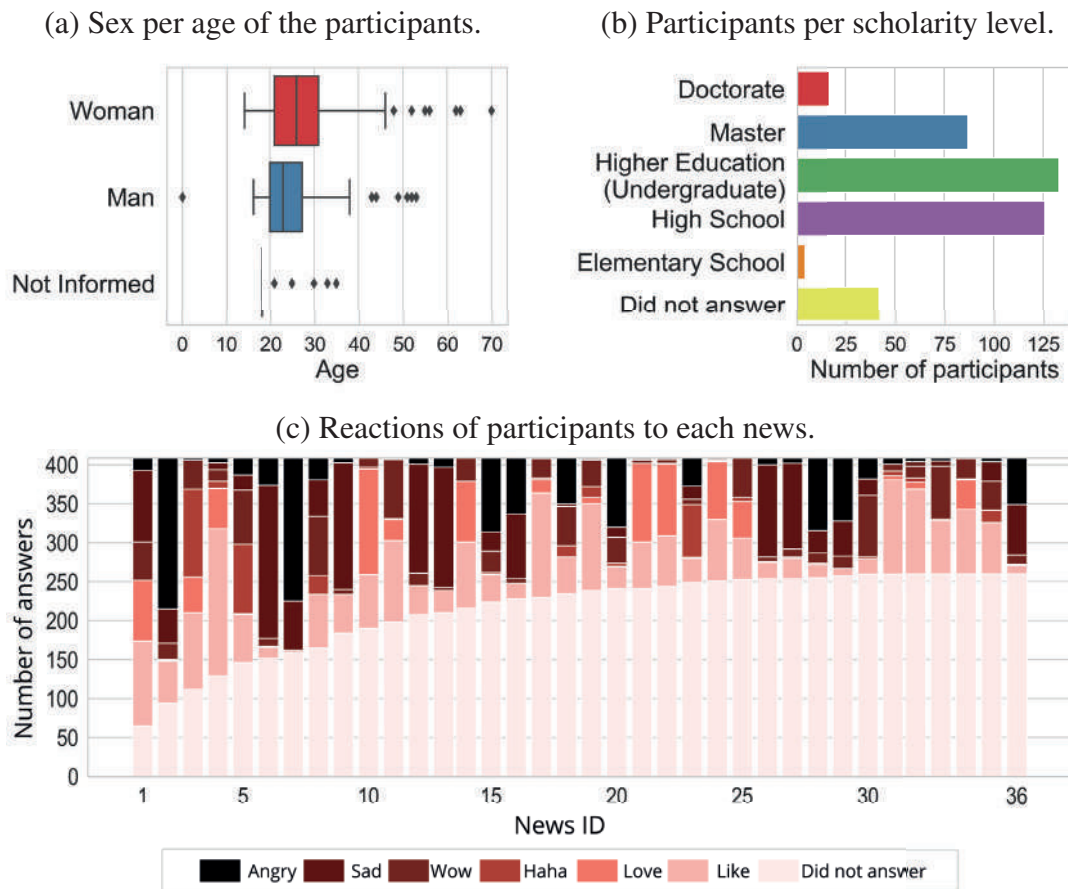
Figure 4. Overview of the participants' information and reactions to the news.

or making decisions based solely on the use of reactions can be challenging. The reactions or emoticons are just a digraph representation, which can involve motivational, excitement and even cultural aspects. For example, in the case of identifying expression of the basic emotions, as explored in [Giuntini et al. 2019] using the set of complete answers of *ReactSet*, it does not seem to be a complex task. However, from these basic emotions, other more sophisticated ones can be extracted. For this, a research effort is required, so that algorithms seek to use context information, or even performing the intersection of basic emotions. A theoretical example is provided by Figure 5 (extracted from [DeMiglio and Williams 2016]), which shows the componential model of emotions defined by [Scherer 1987], along with wellness behavior. Notice that, at the extreme of the axes (between the leaves depicted in the figure), we can see a combination of a set of emotions to define a more sophisticated emotional expression. In addition, data science researches have sought to achieve the same or better performance of techniques commonly applied in Big Data over small datasets [Faraway and Augustin 2018].

4. Download and Citation Request

ReactSet is publicly available for research use, under the Creative Commons license. The dataset is available at *https://github.com/mtcazzolato/reactset*, with the following organization of files:
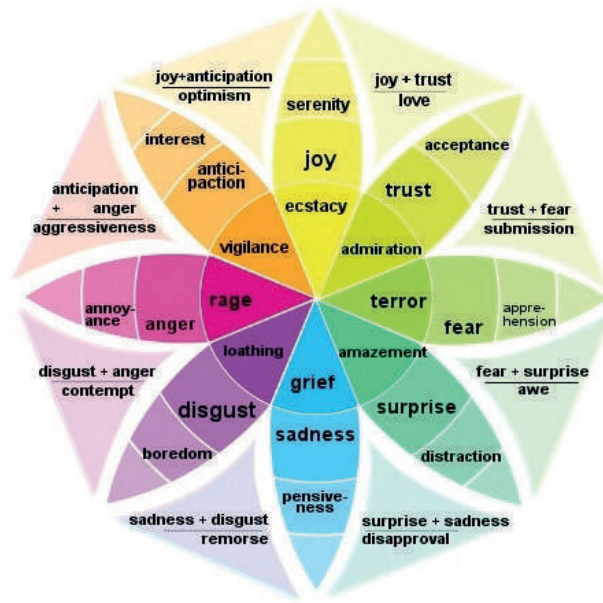
Figure 5. Compositional frame of well-being, based on the componential theory of the emotions of Scherer, extracted from [DeMiglio and Williams 2016]

- *ReactSet/*
    - *newsImages/*: folder with images of extension ".png", each corresponding to a news from table "News" (36 images).
    - *PersonalInfo.csv*: file with data from table "PersonalInfo" (409 rows).
    - *AnswersNews.csv*: file with data from table "AnswerNews" (14,724 rows).
    - *News.csv*: file with data from table "News" (36 rows).
- *README.md*: Read me file with the dataset description and relevant information.
- *Questionnaire.pdf*: A document with the applied questionnaire, with all questions and possible answers.

In case *ReactSet* dataset (or part of it) is used for scientific, industrial, and/or academic purposes, or in case the dataset is publicly mentioned for whatever purposes, we request the acknowledgment of the authors by citing this work.

5. Conclusion

The study of emotions has attracted the attention of the scientific community for decades now. With the advance of social networks, many studies have been proposed to analyze users' behavior regarding different content. In this work, we approached this issue by providing a structured data source, collected from Facebook users. We presented *ReactSet*, a dataset with the reactions of users regarding different news. *ReactSet* gathers personal information from users, such as sex, age, and the number of hours it spends on Facebook. The dataset contains the information of 36 news, along with the number of reactions (Like, Love, Haha, Wow, Sad, Angry) it has received up to July 2019. For each news, we obtained reactions, classifications (positive, neutral, negative), and emotions (anger, disgust, fear, joy, sadness, and surprise) reported by users, considering both the image and the title of the news. We also provide the time the user spent to answer the questions related to each news, and the total time users took to answer the entire questionnaire.

The news selection went through a rigorous process involving postgraduate specialists in the area of Behavior Analysis, in order to search for the most emotionally significant images. The level of agreement among users was also evaluated. It is worth noting that we do not address the news that could be directed to a single, specific audience presenting or ideological or political bias, such as news of famous people or candidates for political offices. Although Facebook was the platform employed for data gathering, the reactions are expressions of the basic emotions defined by Ekman [Ekman et al. 1969, Ekman 1993] as shown in Figure 5. Accordingly, and without loss of generalization, *ReactSet* can be further expanded and explored.

We showed different data visualizations and application challenges to emphasize potential application scenarios for future analysis of *ReactSet*, which includes aspects of temporality, missing data, and recognition of sophisticated feelings and emotions. *ReactSet* is publicly available for download, under the Creative Commons license, and can be downloaded at *https://github.com/mtcazzolato/reactset*. We reinforce that all gathered information was obtained under the approval of the Ethical Committee of our institution, also respecting and preserving the identity of the participants of our study.

Acknowledgments

Ethical Approval

This work has the approval of the Ethics Committee from the School of Arts, Sciences and Humanities (EACH) of the University of São Paulo, Brazil, under Register Number 88799118.8.0000.5390.

References

DeMiglio, L. and Williams, A. (2016). A sense of place, a sense of well-being. In *Sense of place, health and quality of life*, pages 35–50. Routledge.

Ekman, P. (1993). Facial expression and emotion. *American psychologist*, 48(4):384.

Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124.

Ekman, P., Sorenson, E. R., and Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, 164(3875):86–88.

Faraway, J. J. and Augustin, N. H. (2018). When small data beats big data. *Statistics & Probability Letters*, 136:142 – 145. The role of Statistics in the era of big data.

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. Academic Press.

Frijda, N. H., Manstead, A. S., and Bem, S. (2000). *Emotions and beliefs: How feelings influence thoughts*. Cambridge University Press.

Gaspar, R., Pedro, C., Panagiotopoulos, P., and Seibt, B. (2016). Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events. *Computers in Human Behavior*, 56:179 – 191.

Giuntini, F. T., Ruiz, L. P., Kirchner, L. D. F., Passarelli, D. A., dos Reis, M. J. D., Campbell, A. T., and Ueyama, J. (2019). How do I feel? Identifying emotional expressions on facebook reactions using clustering mechanism. *IEEE Access*, 7:53909–53921.

Huang, A. H., Yen, D. C., and Zhang, X. (2008). Exploring the potential effects of emoticons. *Inf. Manage.*, 45(7):466–473.

Jibril, T. A. and Abdullah, M. H. (2013). Relevance of emoticons in computer-mediated communication contexts: An overview. *Asian Social Science*, 9(4):201.

Lerner, J. S. and Keltner, D. (2001). Fear, anger, and risk. *Journal of personality and social psychology*, 81(1):146.

Luor, T. T., ling Wu, L., Lu, H.-P., and Tao, Y.-H. (2010). The effect of emoticons in simplex and complex task-oriented communication: An empirical study of instant messaging. *Computers in Human Behavior*, 26(5):889 – 895.

Oleszkiewicz, A., Karwowski, M., Pisanski, K., Sorokowski, P., Sobrado, B., and Sorokowska, A. (2017). Who uses emoticons? data from 86702 facebook users. *Personality and Individual Differences*, 119:289 – 295.

Russell, J. A., Bachorowski, J.-A., and Fernández-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, 54(1):329–349. PMID: 12415074.

Scherer, K. R. (1987). Toward a dynamic theory of emotion. *Geneva studies in Emotion*, 1:1–96.

Spoor, J. R. and Kelly, J. R. (2004). The evolutionary significance of affect in groups: Communication and group bonding. *Group Processes & Intergroup Relations*, 7(4):398–412.

Statista (2019). Number of monthly active Facebook users worldwide as of 1st quarter 2019 (in millions). https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/. Acessed: 2019-07-7.

Vashisht, G. and Thakur, S. (2014). Facebook as a corpus for emoticons-based sentiment analysis. *Int. J. Emerg. Technol. Advan. Eng*, 4:904–908.

Waller, B. M., Cray Jr, J. J., and Burrows, A. M. (2008). Selection for universal facial emotion. *Emotion*, 8(3):435.

Wang, H. and Castanon, J. A. (2015). Sentiment expression via emoticons on social media. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 2404–2408.

Wegrzyn-Wolska, K., Bougueroua, L., Yu, H., and Zhong, J. (2016). Explore the effects of emoticons on twitter sentiment analysis. *Computer Science & Information Technology*, page 65.

Wolf, K. (2015). Measuring facial expression of emotion. *Dialogues in clinical neuroscience*, 17(4):457.