

# Opinion Label: A Gamified Crowdsourcing System for Sentiment Analysis Annotation

Douglas Cirqueira  
Institute of Technology  
Federal University of Pará  
Belém, PA, Brazil  
douglas.cirqueira@itec.ufpa.br

Lucas Vinícius  
Center of Behavior Theory and  
Research  
Federal University of Pará  
Belém, PA, Brazil  
lvinius123@gmail.com

Márcia Pinheiro  
Institute of Technology  
Federal University of Pará  
Belém, PA, Brazil  
marcia.pinheiro@itec.ufpa.br

Antônio Jacob Junior  
Technology Sciences Center  
State University of Maranhão  
São Luís, MA, Brazil  
antoniojunior@professor.uema.br

Fábio Lobato  
Institute of Engineering and  
Geosciences  
Federal University of Western Pará  
Santarém, PA, Brazil  
fabio.lobato@ufopa.edu.br

Ádamo Santana  
Institute of Technology  
Federal University of Pará  
Belém, PA, Brazil  
adamo@ufpa.br

## ABSTRACT

Online Social Networks are more than a source of large amounts of data, but a potential treasure for companies and institutions that seek information about how they are perceived by their audience. Sentiment Analysis (SA) is a technique that allows the automatic mining of opinions, which can be applied in this context. However, such approach faces many challenges reported in the state of art. Among those, there is the lack of labeled datasets for sentiment classification algorithms. This work presents a platform for SA annotation, with crowdsourcing (CS) and gamification principles, aiming to solve this challenge. Differential aspects of this tool are the possibilities of labeling data for fields related to SA, such as emotion and subjectivity tagging, besides other common tasks in the area, planned to be added in a future version of the tool.

## KEYWORDS

Sentiment Analysis, crowdsourcing, gamification, annotation, text tagging, text mining

## 1 INTRODUCTION

The big social media data represents a new treasure for companies and service providers, which are looking for improving their businesses and products [1]. Many platforms such as forums, blogs, online social networks and news reports contain a sea of opinions and thoughts regarding a diversity of political movements, company & marketing strategies, and services [2]. Therefore, analyses of this scenario can provide new insights regarding the competitive landscape in the market [3].

Sentiment Analysis is the field that can deal with the context previously cited, automatically mining opinions on textual data and classifying usually as positive, negative or neutral [4]. The applications in this field cover, for instance, analysis of product reviews, stock markets, news articles, and even political debates [5].

Although the characteristics of SA are promising, the field faces some challenges in a variety of scenarios, ranging from architecture specific ones [6] to those related to application domains with noisy or scarce datasets [7]. This work highlights the problem of lack of labeled data, also reported in [8], which can pose a barrier in the advancements in this area. For example, when supervised machine learning algorithms are applied for sentiment classification, annotated datasets are needed to train models. Indeed, [9] points out that there are few examples of “ground truth” dataset annotations with human perceptions of social information contained within texts, which could also help on the SA performance reported in the literature.

One road to tackle the previous highlighted issue is the use of crowdsourcing or human collective intelligence methodology [10]. This approach gives the chance to increase workforce for achieving a goal, which can be applied to the task of obtaining labeled data for specific classification tasks [11].

Another important aspect of a crowd system is the compensation policy or incentives, which can help on the performance, as well as the quality of evaluations [12]. Gamification has also been applied as an incentive aspect for CS platforms [13], which can enhance crowd motivation through game scenarios and experiences. However, a CS system specifically developed for SA and related fields, using gamification principles where researchers and professionals are also able to input their own data, has not been found in the literature.

Therefore, this paper has as main goal the proposal of Opinion Label (OL), a gamified crowdsourcing system for the task of SA

and related areas, which aims to help on the lack of annotated corpora problem for the field. Users from academia and industry are then able to input their own data and obtain labeled data back, according to the tasks they have set up for contributors.

## 2 RELATED WORK

There has been a trend to use gamification elements to solve a variety of problems, as an approach to support the enhancement of positive patterns in user's activity into a broad range of systems [14], including crowdsourcing platforms [15]. The main aim is to improve user experience and engagement in those non-game applications and services [16]. For instance, a famous example is [11], which proposed a game for labeling images and was later adopted by Google as an image labeler platform.

The field of SA also has some examples into this direction. To point out, Google has its own mobile gamified CS application<sup>1</sup>, where users can improve Google's services via their contributions. This app also has SA task. Additionally, in [17], the authors defined the Sentiment Quiz, a game where players can vote the polarity of words in a lexicon, ranging from most negative to most positive. In [9], the authors built a platform where users were able to create messages, while others could interpret and input annotations for such data regarding emotions, intentions and attitudes, receiving points for correct answers. Moreover, [18] presents a multiplayer game, where teams must quickly highlight sentiment carrying terms in sentences. Meanwhile, in Guesstiment [19] the focus is on sentiment detection and lexicon construction, through a platform where one user suggests a label, and another, in a guessing role, agrees or not with that. The proposal of [20] focuses on a gamified and language independent platform for sentiment lexicon building, where a pair of players are matched to point out the sentiment in words, piling up blocks similarly to Tetris<sup>2</sup> game. In [21], players have a multifaceted platform, which allows the knowledge inputs regarding polarity, emotions and common-sense concepts. In [22], the authors developed a Web-based game where students were able to classify the polarity and emotion on tweets, and become the "Sentiment detectors" by scoring higher than other player's scores. In [23], the Opinion Marks system allows users to indicate sentiment in words and choose topic phrases or terms in a text.

Regarding the language aspect, only Google's proposal has activities available for Portuguese language, and only for the translation task. Some authors focused on English or another specific language [9, 18, 19, 21, 23], while [17, 20, 22] had different languages or a universal context. Meanwhile, this project aims to be available for both English and Portuguese, which also approaches the gap of such related platforms for SA annotation in Portuguese. In Table 1, this proposal is compared with the related work, concerning system functionalities and possible annotation tasks in the field of SA and related areas.

**Table 1: Opinion Label and Related Work Annotation Tasks**

Aspect	OL	CS <sup>3</sup>	[17]	[9]	[18]	[19]	[20]	[21]	[22]	[23]
User Own Input	X									
Gamification	X	X	X	X	X	X	X	X	X	
Sentiment	X	X	X		X	X	X	X	X	X
Emotions	X			X				X	X	
Subjectivity	X									
Entities	X									
Irony	X									
Keywords/ Key phrases	X		X	X	X	X	X			X
Ambiguity	X									
Context- Dependency	X							X		

From Table 1, it is possible to observe that the majority of proposals in gamification for SA have been focused on annotating sentiment, emotions, keywords & key phrases, and context dependent terms. However, such proposals lack annotations for subjectivity, irony, and ambiguity; which are covered in this paper proposal, although being aspects to be implemented in a future step. Another contribution of this proposal is that it allows users to input their own datasets to retrieve annotations, while such related work did not focus on this feature.

## 3 OPINION LABEL

Opinion Label is a gamified web based crowdsourcing platform, which supports the task of annotating corpora with sentiment, emotions and subjectivity information. Moreover, the tasks of annotating entities, irony, keywords, ambiguity and context-dependency will be implemented in a future step. In practice, researchers and students in SA field will be able to provide and obtain data for their projects, meanwhile professionals and industry will have a tool to help on increasing the accuracy of their machine learning models to, for instance, analyze consumer opinions regarding their brands and services.

### 3.1 Architecture

The proposed system is implemented using PHP 5.6<sup>3</sup>, Bootstrap web framework<sup>4</sup> and MySQL database<sup>5</sup>. It is composed of four major modules, named: Game, Project Management, Profile Management and Data Management.

The Game module is responsible for the management of all the possible annotation tasks, or phases, that a user can face on the system. In every cycle of evaluations, according to the user level, it randomly picks up text instances and tasks to be performed. The Project Management Module collects the information necessary to create a new project on OL. The Profile Management deals with every user information, including their name, score, level, badges, and ranking. The Data Management module communicates directly with the system database, and it takes care of delivering or saving the data necessary for execution of all the other modules; i.e.

<sup>1</sup> <https://crowdsource.google.com/>

<sup>2</sup> <https://tetris.com/>

<sup>3</sup> <http://php.net/>

<sup>4</sup> <http://getbootstrap.com/>

<sup>5</sup> <https://www.mysql.com/>

evaluation cycles, project creation, or profile information consulting.

Therefore, given the simplicity of the system's architecture and application, the Data Management module was abstracted as a Data Access Object (DAO<sup>6</sup>); providing an abstract interface to the system database. In summary, the idea was to build the system on a modular architecture basis, so further improvements and additions to the framework would be simple to perform by others. Fig 1 shows the OL architecture and its modules.

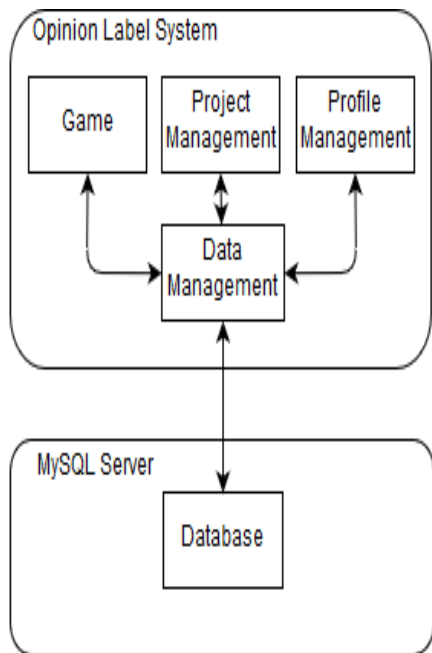


Figure 1: Opinion Label Architecture.

### 3.2 Main Functionalities

Concerning the gamification functionality, there are important elements for implementing gaming experiences, which are usually present in gamified platforms. The work of [24] defines those elements as points, levels, leaderboards, badges, challenges & missions, integration, engagement loops, customization, reinforcement & feedback, and rules. In Table 2, it is possible to notice which gamification elements are applied on OL.

The element of Challenge & Missions is not yet implemented because, in this first stage of development, OL presents a linear sequence of tasks for its users. However, it is an element likely for future consideration. Customization is also out of context so far, given it is not a crucial item for the platform to give its main outcome for users, which are labeled datasets.

Table 2: Gamification Elements on Opinion Label

Element	Applied	Description
Points	Yes	Score for each completed task on the game.
Levels	Yes	Levels with different tasks and complexity.
Leaderboard	Yes	Ranking of players according to their contributions.
Badges	Yes	Medals based on how often users contribute.
Challenges & Missions	No	Extra challenges and missions, besides the main goals and tasks.
Integration	Yes	Explanation of system functionalities to new users.
Engagement Loops	Yes	Call to action and visible demonstration of progress via notifications during the game.
Customization	No	Customization of items on the platform by its users.
Reinforcement & Feedback	Yes	Accessible information to players regarding achievements, score and their current level in the game.
Rules	Yes	Rules to define how to play the game and possible actions inside it.

The first step for users into the OL system is to register into the platform, providing their basic information or via their Facebook account, then being able to go to the three main functionalities into the system, named: Play, Create Project and My Profile.

In order to better illustrate such features, a demonstration video of the platform is available on: <https://youtu.be/psEPBLpxRsE>. This video introduces a tour through the system, providing details regarding its functionalities, as described next.

**3.2.1 Play Functionality.** Here the system will randomly select a batch of random text samples, from different project datasets available. Next, it will select different tasks according to the player current level. For example, on the first level, a user is able to evaluate only sentiment. On further levels, the tasks of emotion detection and subjectivity analysis will be opened, and so on. When users achieve the end of a cycle of evaluations, they will see a notification with the option to play again. In case users do not complete an evaluation cycle, they will not obtain the full points for a completed round, and whenever they return to the system the cycle will restart with new random evaluations. Regarding the user interface, it varies according to the task. For instance, in the sentiment, emotion and subjectivity detection tasks, the user has to choose the option that best fits a text by pressing buttons, representing each available label; that is, for sentiment, the user can choose among positive, negative, neutral, and even strong positive or negative. In the case of entities, irony, keywords, ambiguity and context-dependency tasks, the user will be able to tell if a text has one of those aspects, and tag which terms represent it.

**3.2.2 Create Project Functionality.** This functionality regards the option that users have to create their own projects, providing their own datasets for receiving contributions on OL. In this step, the project author has to provide a project name, description, number of evaluators per instance, main goals (choose, among the available annotation possibilities, which ones are desired to their data be used for, e.g. sentiment and subjectivity detection), source

<sup>6</sup> <http://www.oracle.com/technetwork/java/dataaccessobject-138824.html>

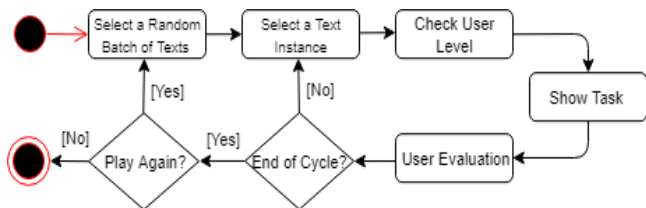
of the data, project end date, and a file with the data. This file has to be structured as one text instance per line. After providing all this information, the project created goes to an evaluation phase. Then, if approved, it is released to receive contributions through the OL user's activity. The author is notified whenever all the project instances get the minimum number of evaluations in at least one of the tasks selected.

3.2.3 *My Profile Functionality*. This component refers to the screen with the own user information, such as name, contributions, and score. The players are able to see information regarding the current badges and achievements obtained, as well as their current position on the OL ranking. The score of each user is based on the number of contributions and on how often they contribute to the system. Some badges are presented in Table 3.

**Table 3: Opinion Label badges**

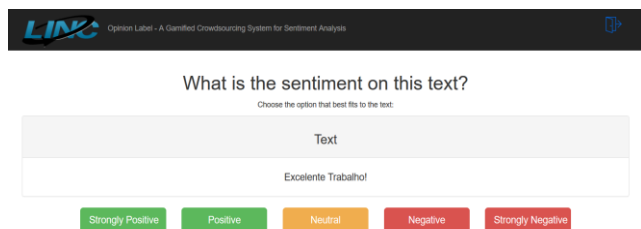
Badge	Description
First Evaluation	First evaluation completed
New Level Reached	When users advance to a new level
First Cycle Completed	Firs cycle of evaluations completed
Number of Contributions	100, 200, 300, 400, 500 contributions
Number of Cycles	5, 10, 15, 20, 25, 30 cycles completed
Champion	First place in ranking
Social Sharing	User shared score on social media

In Fig 2, the activity diagram of the system is highlighted, which corresponds to the moment when a user chooses the Play functionality described previously.



**Figure 2: Activity Diagram from Play functionality.**

In Fig 3, the sentiment task interface is illustrated, with a text instance and the possible labels users can choose for it, as previously described.



**Figure 3: Sentiment Analysis task interface.**

Moreover, Table 4 describes the evaluation tasks available in the Play functionality, and the levels that users have to reach in order to play them.

**Table 4: Evaluation Tasks in Opinion Label**

Task	Level	Description
Sentiment	1	The polarity of a text, such as positive, negative and neutral.
Emotion	2	Emotion on a text, for instance, "joy", "anger", "love".
Subjectivity	3	Whether a text has any opinion.
Entities	4	The sentiment targets or entities in a text.
Keywords	5	Keywords or phrases that help in inferring the sentiment.
Irony	6	Whether a text is ironic and which terms represent the irony.
Ambiguity	7	Whether a text has ambiguity and which terms are ambiguous.
World Knowledge	8	Whether a text needs external or common sense knowledge to be understood, and which entities represent it.

## 4 SYSTEM EVALUATION

A questionnaire with 9 questions regarding user interface and usability of the system was applied for 17 undergrad and master's students in the fields of computer science & engineering, and psychology. The feedback was based on a scale from 1 (very bad), 2 (bad), 3 (regular), 4 (good) to 5 (very good). Table 5 summarizes the aspects highlighted on the survey, and the percentage of votes for each option. The highest percentage is always in bold format.

**Table 5: System Evaluation Results**

Aspect	1	2	3	4	5
Easiness	0%	0%	5.8%	<b>47.1%</b>	<b>47.1%</b>
Intuitiveness	0%	0%	5.9%	29.4%	<b>64.7%</b>
Positions of Text and Buttons	5.9%	5.9%	17.6%	29.4%	<b>41.2%</b>
System Feedback	11.8%	0%	29.4%	17.6%	<b>41.2%</b>
Progress Notifications	6%	0%	17.6%	17.6%	<b>58.8%</b>
Help & Support	5.9%	0%	41.2%	17.6%	<b>35.3%</b>
Colors	5.9%	5.9%	11.7%	29.4%	<b>47.1%</b>
Understanding of Tasks	0%	0%	29.4%	<b>53%</b>	17.6%
Cycle Time	0%	6%	17.6%	23.5%	<b>52.9%</b>

From Table 5, it is noticeable that OL was mostly evaluated as regular, good or very good in all the aspects. However, it is clear the necessity to improve in some features, such as the understanding of the tasks, with more clarifications about how to perform each activity on the platform.

Opinion Label: A Gamified Crowdsourcing System for Sentiment Analysis Annotation

WEBMEDIA'17, October 2017, Gramado, Rio Grande do Sul  
Brazil

## 5 LICENSE AND DISTRIBUTION

Opinion Label will be freely available under GNU GPLv3 license, which is a free, copyleft license for software. The reason for this choice is that it guarantees the freedom to share and change all versions of a program. The platform will be hosted soon at a Github repository for the community to use, test and contribute to the system.

## 6 FINAL REMARKS

This paper proposed a gamified crowdsourcing system with the main goal to provide the research and industry communities in the fields of SA, natural language processing and text mining, a new freely available tool for obtaining annotated datasets; also with a gaming experience to engage more contributions out of the crowd.

One advantage of the OL system, in comparison with the related work, is the ability its users have to input their own data and choose the most interesting annotation tasks to be applied to it.

Moreover, as future work, the tasks of annotating entities, irony, keywords, ambiguity, and context-dependency in texts, which are already described in the system functionalities, are planned for implementation. Another point to be added is the exclusivity of projects, which means a user will be able to share the link to receive evaluations only for his dataset. A dashboard will be implemented, with information concerning the statistics of contributions, agreement coefficient, and an option to export the labeled dataset as a Comma-Separated Value (CSV) file.

Concerning the user interface and gamification elements, some features are planned as well. For example, to show up the ranking on the first screen of the system, so users might be more motivated to start contributing; progress bar during a cycle, to indicate how many evaluations are left to finish; indication of how much experience is left to evolve to a next level; and to review the colors used in the system, according to psychological aspects of sentimental and emotional states, for instance, to use on buttons colors. The social button from Facebook for sharing current score and achievements is also planned, which can help on disseminating the system to new users. User's friends adoption of OL is also another badge to be implemented.

Finally, besides the evaluation already performed, which gave insights on what to improve, another one focused on ergonomic criteria will be performed. The idea is to analyze the platform regarding well-known metrics in the fields of software engineering and human-computer interfaces, by skilled web developers.

## ACKNOWLEDGMENTS

This research was supported by the National Council for the Improvement of Higher Education (CAPES); the National Council for Scientific and Technological Development (CNPq); Amazon Foundation for Studies and Research (Fapespa); and the German Academic Exchange Service (DAAD).

## REFERENCES

- [1] Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82-89.

- [2] Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14-46.
- [3] Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, 55(4), 919-926.
- [4] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- [5] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113.
- [6] Liu, B. (2010). Sentiment analysis: A multi-faceted problem. *IEEE Intelligent Systems*, 25(3), 76-80.
- [7] Balahur, A., & Jacquet, G. (2015). Sentiment analysis meets social media—Challenges and solutions of the field in view of the current information sharing context.
- [8] Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information Fusion*, 36, 10-25.
- [9] Pearl, L., & Steyvers, M. (2010, June). Identifying emotions, intentions, and attitudes in text using a game with a purpose. In *Proceedings of the naacl hlt 2010 workshop on computational approaches to analysis and generation of emotion in text* (pp. 71-79). Association for Computational Linguistics.
- [10] Brabham, D. C. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, 14(1), 75-90.
- [11] Von Ahn, L., & Dabbish, L. (2004, April). Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 319-326). ACM.
- [12] Allahbakhsh, M., Benatallah, B., Ignjatovic, A., Motahari-Nezhad, H. R., Bertino, E., & Dustdar, S. (2013). Quality control in crowdsourcing systems: Issues and directions. *IEEE Internet Computing*, 17(2), 76-81.
- [13] Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does gamification work?—a literature review of empirical studies on gamification. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 3025-3034). IEEE.
- [14] Hamari, J. (2013). Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic commerce research and applications*, 12(4), 236-245.
- [15] Morschheuser, B., Hamari, J., & Koivisto, J. (2016, January). Gamification in crowdsourcing: a review. In *System Sciences (HICSS), 2016 49th Hawaii International Conference on* (pp. 4375-4384). IEEE.
- [16] Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011, May). Gamification. using game-design elements in non-gaming contexts. In *CHI'11 extended abstracts on human factors in computing systems* (pp. 2425-2428). ACM.
- [17] Weichselbraun, A., Gindl, S., & Scharl, A. (2011, October). Using games with a purpose and bootstrapping to create domain-specific sentiment lexicons. In *Proceedings of the 20th ACM international conference on Information and knowledge management* (pp. 1053-1060). ACM.
- [18] Al-Subaihini, A. A., Al-Khalifa, H. S., & Al-Salman, A. S. (2011, December). A proposed sentiment analysis tool for modern arabic using human-based computing. In *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services* (pp. 543-546). ACM.
- [19] Musat Thisone, C. C., Ghasemi, A., & Faltings, B. (2012, July). Sentiment analysis using a novel human computation game. In *Proceedings of the 3rd Workshop on the People's Web Meets NLP: Collaboratively Constructed Semantic Resources and their Applications to NLP* (pp. 1-9). Association for Computational Linguistics.
- [20] Hong, Y., Kwak, H., Baek, Y., & Moon, S. (2013, May). Tower of Babel: a crowdsourcing game building sentiment lexicons for resource-scarce languages. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 549-556). ACM.
- [21] Cambria, E., Xia, Y., & Hussain, A. (2012, May). Affective Common Sense Knowledge Acquisition for Sentiment Analysis. In *LREC* (pp. 3580-3585).
- [22] Furini, M., & Montanero, M. (2016, June). TSentiment: On gamifying Twitter sentiment analysis. In *Computers and Communication (ISCC), 2016 IEEE Symposium on* (pp. 91-96). IEEE.
- [23] Kwon, B. C., Choo, J., Kim, S. H., Keim, D., Park, H., & Yi, J. S. (2015). Opinion Marks: A Human-Based Computation Approach to Instill Structure into Unstructured Text on the Web. In *KDD 2015 Workshop on Interactive Data Exploration and Analytics (IDEA'15)* (pp. 47-55).
- [24] Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. O'Reilly Media, Inc.