Sentimentator : Gamifying Fine-grained Sentiment Annotation

Öhman, Emily Sofi

CEUR Workshop Proceedings
2018

Öhman , E S & Kajava , K S A 2018 , Sentimentator : Gamifying Fine-grained Sentiment Annotation . in E Mäkelä , M Tolonen & J Tuominen (eds) , Digital Humanities in the Nordic Countries 2018 . CEUR Workshop Proceedings , no. 2084 , CEUR Workshop Proceedings , Aachen , pp. 98-110 , Digital humanities in the Nordic Countries DHN2018 , Helsinki , Finland , 07/03/2018 .

http://hdl.handle.net/10138/263706

Sentimentator: Gamifying Fine-grained Sentiment Annotation

Emily Öhman & Kaisla Kajava
University of Helsinki
firstname.lastname@helsinki.fi

February 5, 2018

Abstract

We introduce *Sentimentator*; a publicly available gamified web-based annotation platform for fine-grained sentiment annotation at the sentence-level. *Sentimentator* is unique in that it moves beyond binary classification. We use a ten-dimensional model which allows for the annotation of 51 unique sentiments and emotions. The platform is gamified with a complex scoring system designed to reward users for high quality annotations. *Sentimentator* introduces several unique features that have previously not been available, or at best very limited, for sentiment annotation. In particular, it provides streamlined multi-dimensional annotation optimized for sentence-level annotation of movie subtitles. Because the platform is publicly available it will benefit anyone and everyone interested in fine-grained sentiment analysis and emotion detection, as well as annotation of other datasets.

1 Introduction

The main problem with, even conventional, sentiment analysis methods tends to boil down to a lack of tagged corpora. Proper annotation is costly and can be unfeasible in some cases [6]. *Sentimentator* addresses this lack of annotated corpora, and provides a novel tool for producing datasets efficiently that cover a wide range of genres (within the domain of movie subtitles).

A crowd-sourced gamified annotation scheme based on Plutchik's eight emotions [25] as well as the sentiments of positive, negative, and neutral presents new opportunities, but also challenges. It is more time consuming and requires more reflection on the part of the annotator to tag a sentence with more than two or three dimensions. We solve this by gamifying the process in order to (1) have a simple and straightforward user interface for the annotation, and (2) present an inviting option for students and other non-experts to help with the annotation by setting up a game-like platform.

The reason we have chosen to gamify the annotation process is the increased accuracy [22] and lower price compared to more traditional crowd-sourcing methods.

We want to produce more training data easily with lower cost to train better machine learning-based classifiers on top of the annotated datasets.

The output of sentiment analysis is often expressed as a numeric value on a sliding scale of negative, neutral, and positive sentiments or simply a ternary score of one of the aforementioned values. This approach is limited [4] and applicable only to some of the myriads of possible uses for sentiment analysis.

For this to be feasible, a new approach beyond positive and negative is necessary. We propose to use Plutchik's eight core emotions [25] (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) alongside the sentiments of positive and negative typically used in sentiment analysis.

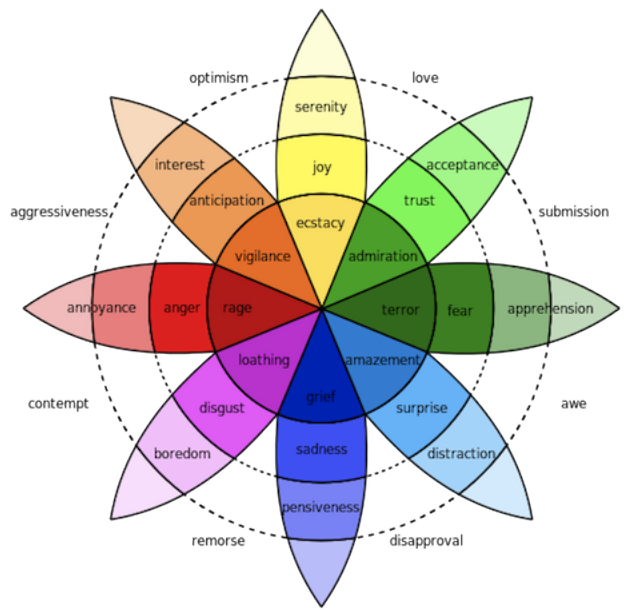

Figure 1: Plutchik's wheel of emotions [1]

With the use of an intensity measure, *Sentimentator* effectively allows for sentiment annotations on the entire wheel. Furthermore, because intensity adjustment and combination of emotions is made possible, the difficulty of the annotation task does not increase linearly with the number of dimensions in our scheme. A further 24 combinations of emotions are possible through combinations of the eight core emotions such that, for example, 'awe' can be expressed through annotating for 'fear' and 'surprise'. Therefore 51 unique emotions and sentiments are described by the *Sentimentator* annotation scheme.

[1]Source of figure and table: https://en.wikipedia.org/wiki/Contrasting_and_categorization_of_emotions

Table 1: Emotions and Opposites[2]

Mild emotion	Mild opposite	Basic emotion	Basic opposite	Intense emotion	Intense opposite
Serenity	Pensiveness	Joy	Sadness	Ecstasy	Grief
Acceptance	Boredom	Trust	Disgust	Admiration	Loathing
Apprehension	Annoyance	Fear	Anger	Terror	Rage
Distraction	Interest	Surprise	Anticipation	Amazement	Vigilance

Although this dataset is only being developed at the moment, when it has been completed and tested, it will be made publicly available. Once we have some annotated data, it can be used as training and testing data. Previous work suggests that our approach, if implemented correctly, should be on par with or better than some of the best methods available at the moment [11, 28]. Fine-grained sentiment analysis provides exciting new avenues of research. With a properly tagged dataset, many researchers will be able to improve the output of their previous methods as it is hard to come by labeled data for sentiment analysis, especially fine-grained [30].

There are a number of areas where sentiment analysis could become an invaluable tool for digital humanities scholars. Some examples of these areas are history, literature, translation studies, language, and social sciences. Possible approaches for historians and social scientists could be to study how the attitude towards a specific topic has changed through time [27]. In literature, story arcs could be analyzed automatically to find over-arching themes and to identify how stories develop within different genres [15, 16], and sociolinguists or translation studies researchers could compare how different languages express emotion and sentiment in what are supposedly identical texts using sentiment analysis on parallel corpora [23].

In section 2 we present an overview of relevant related work and current approaches. In section 3 we discuss gamification from a theoretical perspective and in section 3.1 we shine the light on our framework and platform. In this section we also consider the practical applications of the ideas discussed in section 3 in greater detail. The last two sections are reserved for future work and a concluding discussion.

2 Related Work

There are many approaches to mining data for sentiments. They range from purely lexical to fully unsupervised [14] with many hybrid methods in-between. Andreevskaia et al. [1] suggest that the reason for the prevalence of unsupervised knowledge-based methods in binary sentence classification is the lack of labeled training data. This is the main issue *Sentimentator* will address.

There are a few applications that offer similar solutions to ours on some level (see for example [1, 7, 22, 21, 16]), but none of these are all three: (1) domain-independent, (2) sentence-level annotations, (3) beyond positive and negative, i.e. multi-dimensional or fine-grained.

Most current approaches still focus on the positive-negative axle of polarity. This binary, or at best ternary with 'neutral', approach is far too restricted for many applications [17], and new methods increasingly incorporate other dimensions into sentiment analysis beyond the binary approach. For example Honkela et al. [12] use a five-dimensional PERMA-model (Positive emotion (P), Engagement (E), Relationships (R), Meaning (M) and Achievement (A), and EmoTwitter [21] utilizes a ten-dimensional model (positive, negative, joy, sadness, anger, anticipation, trust, disgust, fear, and surprise) based on the NRC lexicon [20] which in turn uses Plutchik's wheel of emotions.

Although sentence or phrase level sentiment analysis is important for many applications [24], e.g. question and answering tasks [31], there are few sentence-level annotated datasets because of the time-consuming annotation process. There is also a lack of sentiment clues in sentences and other short text spans. If there is only one sentiment clue in a sentence, the entire analysis rests on possibly a single word. Therefore it can be challenging to reach a correct analysis [2]. Wilson et al. [30, 31] show that for sentence-level sentiment analysis to work, it is important to be able to tell when a sentence is neutral. This reduces the risk of assigning sentiments and emotions where there are none and allows for contextually accurate sentiment and emotion analysis. The annotation scheme of *Sentimentator* allows for neutral tagging increasing the likelihood of correct contextual analysis.

There has long been a discussion on how classifiers trained on data from one domain might not work as well when applied on data from a different domain [3, 24, 11]. Therefore Boland et al. [2] suggest annotating training data without context and at sentence-level. Furthermore, ignoring context means that although a sentence is implicitly negative because it is expected that the following sentence is explicitly negative, it should be tagged as positive or neutral (depending on the sentiments in that sentence alone) as otherwise that one sentiment would be weighted double [2]. Our annotation scheme also allows for all possible permutations and mixed sets of the ten dimensions, so there is no issue with mixed sentiments or emotions in a sentence as all can co-exist.

3 Gamifying Annotation

Gamification happens when game elements are used in a non-game context to "improve user experience and engagement" [5]. In the latter half of the 2010s there has been an increase in gamification [9], mainly for marketing [10], but also scientific purposes [8].

Sentimentator players (annotators) select emotions and sentiments for random sentences. Other common gamification elements included in *Sentimentator* are badges, leaderboards, levels/rank, as well as avatars, and feedback. Variation in the gaming content is key to minimize the repetitiveness of tasks. We offer annotators simple annotation tasks, correction of automatically annotated data tasks, and ranking of sentence tasks.

Groh discusses some pitfalls of gamification stating that "pleasure is not additive and rewards can backfire" [9]. We follow the principles described by Deterding [5] and Schell [26] in order to avoid these pitfalls. These principles are (1) Relatedness (connected to other players), (2) Competence (mastering the game problems), and (3)

Autonomy (control of own life).

A simple way to increase the relatedness of our platform is to allow players to see their own and their peers' progress as well as in real-time see how their work impacts their grade (if annotation is part of coursework) or some other real-world benefit. This can be done partly by leaderboards, but also by showing the student a progress bar that shows how close they are to the next goal/rank/level. As with Zooniverse[3] [8] there is an opportunity to be part of a larger scientific community and contribute to the advancement of science, however small the increment. PlanetHunters (Zooniverse) have even offered co-author credits to those who have helped locate new exoplanets via their gamified data analysis platform[4].

For *Sentimentator* to allow annotators to feel competent and that they are improving they need feedback on their progress in relation to others. It is not desirable for the annotators to see how other annotators have annotated the same data, but annotations can still be compared and scored. When comparing annotations with those made by other annotators, the reliability/accuracy score is dependent on the reliability rating of the other annotator. If annotations correlate better with annotators with a higher reliability rating then the score given is also higher and vice versa. This means that the rank of a player also affects how other players score. Additionally, a score is affected by how well the annotation correlates with validated test sentences. See sections 3.1 and 3.2 for more details about gameplay and scoring.

The validated sentences are sentences that have been annotated by expert annotators who have received thorough instructions on how to annotate with the aim of consistency across annotators. The results of these expert annotators will be reviewed before they are used as seed sentences. The "gamer" annotators will receive a similar tutorial via *Sentimentator*, but their annotations will generally only be compared against the validated seed sentences and the annotations of their peers.

The first players of *Sentimentator* are students of language technology. It is difficult to not offer these students extrinsic rewards (in the form of extra credit and such), especially in the initial stages of gathering testing and training data. Some of this loss of autonomy is combated by emphasizing the scientific contribution that they make and keeping them posted about e.g. articles published using datasets they helped create. Once the platform is open to all, however, there is significant autonomy.

3.1 Gameplay

The annotators are greeted by an info screen where they are presented with Plutchik's [25] wheel (see figure 1). They are told how to tag the different emotions (i.e. the emotion of 'remorse' would suggest 'disgust' and 'sadness' of a higher intensity). There are three different ways to play the game. The first one is to get pre-analyzed (tagged by lexical lookup) sentences and adjust the annotation, the second is to get un-tagged sentences and annotate them, and the third is a sentence intensity ranking task.

The first type of gameplay consists of annotating unvalidated pre-annotated sentences. The sentences have been tagged by using simple lexical comparison. The

[3] https://www.zooniverse.org/
[4] https://www.planethunters.org/

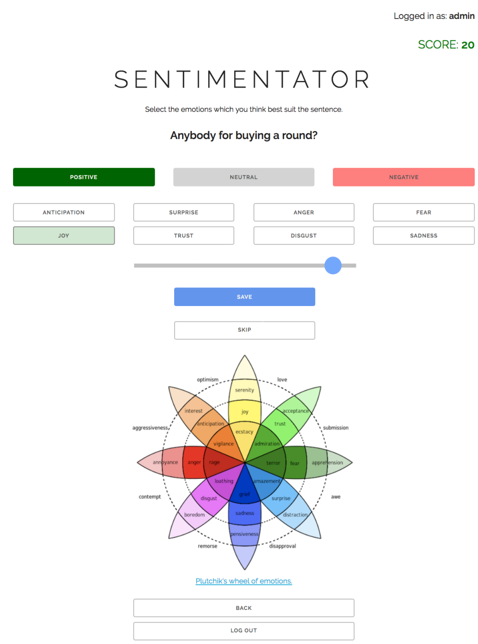

Figure 2: Sentimentator Prototype Interface[5]

annotator/player needs to judge whether the analysis is correct or needs adjustment. The scoring is a simple fraction of full scores until the annotations can be compared to peer annotations.

Both validated and unvalidated sentences are presented to the annotator who does not know which type is in question. The annotator/player needs to recognize the emotions and sentiments present in the sentence without context. The scoring is a simple fraction of full scores until the annotations can be compared to peer annotations. In the case of validated sentences, the scores received follow the formula in 3.2 and significantly impact rank.

All of the gametypes have intensity of sentiments/emotions built into the annotation through the use of a slider that is pre-set to 50%. The slider can be adjusted higher or lower to signify the intensity the annotator judges the sentence to possess. For the ranking task sentences, whether the intensity has been adjusted or not, are shown and by dragging and dropping the sentences in order from most intense to least intense, we are able to get more accurate intensity scores through this best-worst scaling approach [19]. We will use both sentences that have been annotated and those that have not for this task to get data on how the nature of the task affects intensity rankings.

All annotations are done without giving any context as suggested by the results achieved by Boland et al. [2]. Their research shows that to achieve more accurate

[5]For the prototype css we used: `http://getskeleton.com`

results when using an annotated corpus for training and testing, context is confusing and gives erroneous annotations. The issues with choosing the correct annotation is discussed in section 3.3.

3.2 Scoring

As discussed in section 3 about Gamification, it is important for players to feel competent and like they are mastering a skill. Therefore scoring is one of the most important aspects of gamification. Players need to feel that they are being compensated appropriately for the work they are doing even if it is a game and the compensation is in the form of points.

Players accumulate both rank (R) and level where rank is a prestige or reliability score based on how well the player's annotations correlate with validated test sentences.

$$R = \frac{T_{sv}}{V_{max}}$$

where T_{sv} stands for total score from validated sentences and V_{max} for maximum possible score for the player in question from validated sentences

Level is a straight-forward measure of the number of annotated sentences.

$$Level = \frac{T_{sv}}{V_{max}} \times \frac{A_p}{\frac{T_a}{100}}$$

where A_p stands for total sentences annotated by player and T_a stands for the total number of sentences in the dataset and $0 \leq R \leq 1$, and $0 \leq Level \leq 100$

There are two main types of scores; those based on rank (i.e. prestige or reliability) and those based on validated sentences. All tasks yield a pre-adjustment score (S). This score is the score that is based on simply doing the task without any regard to how well the task has been completed or how it correlates to other players annotations.

The calculations for the score received from annotating validated sentences (S_v) is fairly straight-forward.

$$S_v = \frac{S}{V_s}$$

where V_s stands for the max score possible for that task as per the score for the validated sentence

As for the score based on peer annotation (S_p), this score accumulates rank only after a certain number of annotations have been made for the same sentence. The rank (or reliability/prestige) rating of the annotator (R_{oa}) who has annotated the same sentence before influences the score for the annotation as per the following:

$$S_p = \frac{P_s}{S_{oa}} \times R_{oa}$$

where P_s stands for the pre-adjustment annotation score of the peer and S_{oa} stands for the score of the other annotator. In practice this will work much like a weighted average across all peers. The number of annotators per sentence is also limited.

The rank influences the score as it is at the time of the annotation, i.e. the rank that was valid at the time of the annotation is considered. If an annotators rank improves or declines, it is a reflection of their annotation skill in real-time, not when they did the original annotation. Therefore dynamic scoring would not accurately reflect the reliability of an annotation.

3.3 Choosing the Right Annotation

There is a lot to consider when choosing the right annotation. It is virtually impossible for all annotators to annotate every sentence exactly the same. This results in noisy annotations. Hsueh et al. [13] discuss measures to control the quality of annotations. In their study they compare noisy annotations against the gold standard labels. As we do not have the option to compare to a gold standard, we will have to rely heavily on the scores received for annotating validated sentences (see Scoring). However, with enough annotations we will be able to remove annotations made by the noisiest group of annotators (In Hsueh et al. [13] this group consisted of 20% of annotators).

As our scoring already relies on validated sentences even when annotating unvalidated sentences, we are unlikely to need much screening for noisy annotations. It is, however, important to keep the possibility of excluding noisy annotators from the final annotation output. It is also important to be able to exclude ambiguous examples from the annotations in order to maximize the quality of the labels [13]. Even though this is an issue for after we have annotated data, it is an important aspect to keep in mind when creating the framework.

All annotations will have been annotated by at least three annotators before they are made final. Naturally, these tags will not always be identical. The way *Sentimentator* is constructed allows for easy checking of differing annotations. The first step is the automatic comparison against validated sentences. The second is to defer to the annotation made by the highest ranked annotator. However, where discrepancies are deemed considerable, annotations can be flagged to be reviewed by experts.

3.4 Data

We use the publicly available dataset OPUS.[6] Our initial focus is the English and Finnish parallel corpus of movie subtitles, but the number of possible languages to annotate is only limited by the data itself. The current version has been tested on eight languages. We chose movie subtitles [29, 18] as they contain a lot of emotional content in a style applicable to many different types of tasks [23], and because a high-quality parallel corpus exists for many different languages.

[6]http://opus.lingfil.uu.se - We use the newest, 2018, version which has at the time of writing not yet been made publicly available.

4 Future Work

The evaluation of this framework can only begin once a certain amount of lines have been annotated and cross-checked. For a demonstration, some results can be achieved with approximately 1000 lines annotated, but for proper sentiment analysis at least four times that is required. This means that at least three people will need to annotate 4000 lines, preferably many more people annotating tens of thousands of lines/sentences.

One simple way of spreading out this task, and to be able to utilize expert annotators for a low cost, is to outsource it as extra-credit coursework in computational linguistics, or corpus linguistics courses and similar. Once enough data has been annotated for training and testing data, we can evaluate our framework and compare it against the current gold-standard.

We plan on evaluating the final dataset by taking into account both the distribution of the data and classification performance using a set of different classifier types. We intend to evaluate the distributional balance of the data in regard to the amount and quality of lines/sentences of each label or label combination. This way we reveal patterns in the dataset which may affect classification results. For example, sentences of a given label may be considerably longer or shorter than sentences of another label, or contain rare words. Similarly, the sentences may originate from a movie of a specific genre or time period and thus contain a particular type of language use, such as jargon or archaic words. This allows us to evaluate the sparsity of the data in both the dataset as a whole as well as across different labels. We can then assess whether some parts of the dataset are more sparse and thus less likely to allow classifiers to detect meaningful patterns.

Using a set of different classifiers also allows us to evaluate the quality of the dataset. By building confusion matrices for each classifier, we can observe the classification accuracy, precision, recall, and F-measure for each class in the dataset as well as the overall performance of the classifier.

Other future work includes testing the finalized semi-supervised algorithm on actual datasets. In addition to the suggestions in the Introduction, some possible explorations could be newspaper or online discussion forum data dumps with the search keys for migration and other current issues.

A comprehensive set of high-quality annotations also allows for comparison between intra-lingual annotations of the same sentences by different users as well as identifying possible patterns in cross-lingual annotations of parallel sentences. Another interesting question to investigate is whether showing users sentences which have already been annotated influences their choices when choosing the most suitable tags for those sentences. In this research setting, users would choose the gameplay option where they evaluate annotated sentences with the task of either accepting or editing those annotations. This data would then be compared to parallel annotations of sentences which users have annotated from scratch.

We also hope that other researchers in various fields including computational linguistics as well as humanities etc. will find both the annotation platform and the dataset useful and publish their own research based on our work.

5 Conclusions and Discussion

We have introduced *Sentimentator*, a publicly available, gamified, web-based annotation tool specifically for fine-grained sentiment analysis. Not only do we go beyond binary sentiment classification, but our annotation scheme allows us even more detailed fine-grained annotation by adjusting the intensity of Plutchik's eight core emotions. The expansion gives us a possible eight core emotions with three intensities each, and 24 combinations of the core emotions with a total of 48 separate emotions and an additional two sentiments plus neutral, i.e. 51 total sentiments and emotions available for annotation (See figure 1 and table 1 for specifics).

The gamification of annotation decreases the cost of annotation and increases the size of the final dataset. It has also been shown to give more accurate annotations than traditional crowd-sourcing methods [2]. Furthermore, we have carefully designed the scoring to reward more accurate annotations and improve the annotation experience by making it more interesting. After initial evaluation tasks, the dataset as well as the platform itself, will be made open to anyone who needs a sentiment annotated dataset.

This type of data is rare to come by, and we have high hopes for the applications of the dataset and the platform itself.

References

[1] ANDREEVSKAIA, A., AND BERGLER, S. Clac and clac-nb: Knowledge-based and corpus-based approaches to sentiment tagging. In *Proceedings of the 4th International Workshop on Semantic Evaluations* (Stroudsburg, PA, USA, 2007), SemEval '07, Association for Computational Linguistics, pp. 117–120.

[2] BOLAND, K., WIRA-ALAM, A., AND MESSERSCHMIDT, R. Creating an annotated corpus for sentiment analysis of german product reviews.

[3] DAVE, K., LAWRENCE, S., AND PENNOCK, D. M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th International Conference on World Wide Web* (New York, NY, USA, 2003), WWW '03, ACM, pp. 519–528.

[4] DE ALBORNOZ, J. C., PLAZA, L., AND GERVÁS, P. Sentisense: An easily scalable concept-based affective lexicon for sentiment analysis.

[5] DETERDING, S., SICART, M., NACKE, L., O'HARA, K., AND DIXON, D. Gamification. using game-design elements in non-gaming contexts. In *CHI'11 extended abstracts on human factors in computing systems* (2011), ACM, pp. 2425–2428.

[6] EHRMANN, M., TURCHI, M., AND STEINBERGER, R. Building a multilingual named entity-annotated corpus using annotation projection. *RECENT ADVANCES IN* (2011), 118.

[7] ERYIGIT, G., CETIN, F. S., YANIK, M., TEMEL, T., AND ÇIÇEKLI, I. Turksent: A sentiment annotation tool for social media. In *LAW@ ACL* (2013), pp. 131–134.

[8] GREENHILL, A., HOLMES, K., LINTOTT, C., SIMMONS, B., MASTERS, K., COX, J., AND GRAHAM, G. Playing with science: Gamised aspects of gamification found on the online citizen science project-zooniverse. In *GAMEON'2014* (2014), EUROSIS.

[9] GROH, F. Gamification: State of the art definition and utilization. *Institute of Media Informatics Ulm University 39* (2012).

[10] HAMARI, J., AND KOIVISTO, J. Social motivations to use gamification: An empirical study of gamifying exercise. In *ECIS* (2013), p. 105.

[11] HE, Y., AND ZHOU, D. Self-training from labeled features for sentiment analysis. *Information Processing & Management 47*, 4 (2011), 606 – 616.

[12] HONKELA, T., KORHONEN, J., LAGUS, K., AND SAARINEN, E. Five-dimensional sentiment analysis of corpora, documents and words. In *Advances in Self-Organizing Maps and Learning Vector Quantization - Proceedings of the 10th International Workshop, WSOM 2014* (2014), pp. 209–218.

[13] HSUEH, P.-Y., MELVILLE, P., AND SINDHWANI, V. Data quality from crowdsourcing: A study of annotation selection criteria. In *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing* (Stroudsburg, PA, USA, 2009), HLT '09, Association for Computational Linguistics, pp. 27–35.

[14] HU, X., TANG, J., GAO, H., AND LIU, H. Unsupervised sentiment analysis with emotional signals. In *Proceedings of the 22nd international conference on World Wide Web* (2013), ACM, pp. 607–618.

[15] JOCKERS, M. L. *Text analysis with R for students of literature*. Springer, 2014.

[16] KAKKONEN, T., AND KAKKONEN, G. G. Sentiprofiler: creating comparable visual profiles of sentimental content in texts. *Language Technologies for Digital Humanities and Cultural Heritage 62* (2011), 189–204.

[17] LI, J., AND HOVY, E. Reflections on sentiment/opinion analysis. In *A Practical Guide to Sentiment Analysis*. Springer, 2017, pp. 41–59.

[18] LISON, P., AND TIEDEMANN, J. Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles. In *LREC* (2016), N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, Eds., European Language Resources Association (ELRA).

[19] MOHAMMAD, S. M., AND BRAVO-MARQUEZ, F. Emotion intensities in tweets. *CoRR abs/1708.03696* (2017).

[20] MOHAMMAD, S. M., AND TURNEY, P. D. Crowdsourcing a word-emotion association lexicon. 436–465.

[21] MUNEZERO, M., MONTERO, C. S., MOZGOVOY, M., AND SUTINEN, E. Emotwitter - a fine-grained visualization system for identifying enduring sentiments in tweets. In *CICLing (2)* (2015), A. F. Gelbukh, Ed., vol. 9042 of *Lecture Notes in Computer Science*, Springer, pp. 78–91.

[22] MUSAT, C.-C., GHASEMI, A., AND FALTINGS, B. 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* (Stroudsburg, PA, USA, 2012), Association for Computational Linguistics, pp. 1–9.

[23] ÖHMAN, E., HONKELA, T., AND TIEDEMANN, J. The challenges of multi-dimensional sentiment analysis across languages. *PEOPLES 2016* (2016), 138.

[24] PANG, B., AND LEE, L. Opinion mining and sentiment analysis. *Information Retrieval 2*, 1-2 (2008), 1–135.

[25] PLUTCHIK, R. A general psychoevolutionary theory of emotion. *Theories of emotion 1* (1980), 3–31.

[26] SCHELL, J. The pleasure revolution: Why games will lead the way, googletechtalks std. november 2011, 2015.

[27] SPRUGNOLI, R., TONELLI, S., MARCHETTI, A., AND MORETTI, G. Towards sentiment analysis for historical texts. *Digital Scholarship in the Humanities 31*, 4 (2016), 762–772.

[28] TÄCKSTRÖM, O., AND MCDONALD, R. Semi-supervised latent variable models for sentence-level sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2* (2011), Association for Computational Linguistics, pp. 569–574.

[29] TIEDEMANN, J. Parallel data, tools and interfaces in opus. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)* (Istanbul, Turkey, may 2012), N. C. C. Chair), K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odijk, and S. Piperidis, Eds., European Language Resources Association (ELRA).

[30] WILSON, T., WIEBE, J., AND HOFFMANN, P. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (2005), Association for Computational Linguistics, pp. 347–354.

[31] WILSON, T., WIEBE, J., AND HOFFMANN, P. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics 35*, 3 (2009), 399–433.