

Comparing Live Sentiment Annotation of Movies via Arduino and a Slider with Textual Annotation of Subtitles

Thomas Schmidt, Isabella Engl, David Halbhuber and Christian Wolff

Media Informatics Group, University of Regensburg, Germany
{firstname.lastname}@ur.de

Abstract. Movies in Digital Humanities are often enriched with information by annotating the text e.g. via subtitles. However, we hypothesize that the missing presentation of the multimedia content is disadvantageous for certain annotation types like sentiment annotation. We claim that performing the annotation live during the viewing of the movie is beneficial for the annotation process. We present and evaluate the first version of a novel approach and prototype to perform live sentiment annotation of movies while watching them. The prototype consists of an *Arduino* microcontroller and a potentiometer which is paired with a slider. We perform an annotation study for five movies receiving sentiment annotations from three annotators each, once via live annotation and once via traditional subtitle annotation to compare the approaches. While the agreement among annotators increases slightly by using live sentiment annotation, the overall experience and subjective effort measured by quantitative post questionnaires improves significantly. The qualitative analysis of post annotation interviews validates these findings.

Keywords: Sentiment Annotation, Sentiment Analysis, Movies, Film Studies, Arduino, Annotation.

1 Introduction

Annotation is an important task in Digital Humanities (DH) in order to enrich cultural artefacts with additional information. While there are annotation tasks that can be carried out automatically, human annotations are still necessary for many DH projects. Various forms of syntactic (cf. [4]) or semantic annotations [3, 6, 41] exist for various media types. In film studies various approaches towards annotation are used. Scholars use annotation tools to annotate information like shot types and lengths [51], camera angles, important movements or people [16] to add a more objective and in-depth understanding of movies accompanying the hermeneutical approach towards interpretation. Film archives employ methods of crowdsourcing to gather metadata information about their movie inventory [37]. Annotations can also be used as data to train and evaluate modern machine learning approaches which have become more and more important in DH in recent years [cf. 8, 15]. One research branch in DH explores the development and evaluation of tools for these various annotation processes [7, 17, 31, 45, 48] and the influence of context, material and task on annotation quality (cf. [28, 32]). The work presented here is in line with this strand of research. As our annotation use case we investigate sentiment annotation in movies.

Please cite as:

Schmidt, T., Engl, I., Halbhuber, D. & Wolff, C. (2020). Comparing Live Sentiment Annotation of Movies via Arduino and a Slider with Textual Annotation of Subtitles. In *Post-Proceedings of the 5th Conference Digital Humanities in the Nordic Countries (DHN 2020)* (pp. 212-223).

Sentiment analysis is the research area concerned with the computational analysis of sentiments, predominantly in written text [25, p. 1]. Sentiment and emotion analysis have been explored in the context of the DH. Researchers explore sentiment analysis in various literary genres like plays [27, 30, 39, 38, 40, 52], novels [20, 23, 36], fairy tales [1, 2] and fan fictions [21, 35] but also in the social media context [44, 46]. While the focus of research is predominantly on text, esp. traditional text genres like novels and plays, research for movies is rare: Öhman and Kajava [33] have developed *Sentimentator*, an annotation tool specifically designed to annotate sentiment and emotion for movie subtitles with gamification elements. They apply the tool to acquire emotion annotated subtitles [19, 34]. Chu and Roy [9] explore multimodal sentiment analysis in videos and focus on short web videos to identify emotional arcs. Schmidt et al. [43] explore multimodal sentiment analysis on theatre recordings with mixed results.

As with most classification tasks, well-curated corpora are an important resource to develop modern machine learning algorithms. However, specifically for the research area of narrative media and texts one can identify a lack of such corpora (cf. [22]). Reasons for this might be that annotators perceive the task as challenging and tedious [1, 41, 42, 49]. If the annotators have no expertise, they report problems with the language and the missing context [1, 41, 42, 49]. Furthermore, narrative texts are generally more prone to subjectivity since they can be interpreted in different ways. Therefore, annotation agreements are typically rather low [1, 41, 42, 49], which is also a problem for the successful creation of corpora. In the context of movies, sentiment or emotion annotation projects are rare and mostly focused on the annotation of the textual content of movies like the subtitles [19, 33]. Similar to literary studies one can identify an interest in more sophisticated concepts besides sentiment like differentiated emotion categories and scales [32]. We will focus on sentiment for our study. While the sentiment concept cannot fully represent the complex emotional expressions in movies, we regard it as simpler and therefore more fitting for this first pilot study. We present a live sentiment annotation solution enabling the annotation of movies while watching them. We argue that this approach is beneficial to more traditional approaches like the annotation of subtitles when dealing with movies.

First, movies are multimedia artefacts and the lack of the presentation of the video channel leads to information loss. Many emotions are expressed via the face and the voice of the actor (or via additional aspects like music, colors, and camera perspectives) and not just the text. Therefore, viewers might be able to annotate sentiment and emotions easier and more consistent when experiencing the entire movie. Additionally, a lot of context that might be important to understand the feelings of the characters might be expressed via other channels than the text. Furthermore, emotions are also often expressed without saying anything in a movie. Textual annotation only allows the annotation of parts in which characters talk, everything else is neglected. While there are video annotation tools that offer the video and audio channel to be used for movie annotation, they often need training before usage and rather support asynchronous work needing to constantly pause and adjust the time and frame of the movie for the annotation [10, 26]. We assume that live annotation during the viewing of the movie facilitates the annotation process because the viewer/annotator can di-

rectly and immediately assign their annotations based on what they are experiencing. Furthermore, the usage of a continuous slider as in our setting might also resemble the rather vague concept of sentiment much more than nominal class assignments [5, 39] or ordinal ratings [29, 50]. Following Nobel laureate Daniel Kahneman’s line of thought, annotating in the actual movie watching situation might come closer to the emotional reality than more reflective post-hoc annotation [18]. Please note that while we focus on sentiment annotation for our study, the system can be similarly used for every sort of emotion or other scale for which one desires live video annotations.

2 Live Sentiment Annotation Approach

2.1 Technical Setup

The annotation system consists of an Arduino microcontroller connected to a linear potentiometer, which is paired with a slider. The Arduino itself is connected to a computer running a Python script. The script represents the core of the system, it is responsible for reading the current value of the slider, logging it and presenting it to the user in a small GUI while watching the movie on a TV. The slider depicts continuously changing resistance levels between 0 and 1023; these values may be translated programmatically to other scales. The python script, running in the background (e.g. on a laptop connected to the TV), records these values simultaneously and it shows the user the current slider position and thus the currently selected value in a simple GUI. Figure 1 depicts the user view for an exemplary application in a movie annotation.



Fig. 1. Example scene from a TV show (left). Python script displaying the currently chosen value on the Arduino slider. The GUI also depicts a rudimentary scale (right).

2.2 Annotation Process

For the annotation process, the annotator/viewer will be presented with the movie and the interface. Additionally, the annotators are equipped with the cased Arduino slider. Figure 2 shows an early prototype of the slider’s encasing.



Fig. 2. Prototype of the slider casing. The shell also features a rudimentary scale allowing the user to navigate without the GUI. Please note that the slider is continuous and not nominal.

The slider will be operated by the user while watching a movie or TV show and is placed at the side of his chair so the viewer can adjust the scale intuitively with his hands, while watching the movie. The slider is portable and can be placed as the viewer wishes. During the study, the value of the Arduino slider, time-stamped, is read and logged by the Python script every 100ms. By saving the timestamp, the slider value can be exactly assigned to a certain time in a film or TV program in a subsequent data analysis. The movie shown as well as the slider are synchronized via a Python script connecting the slider and VLC-player, which is the media player we use to present the movie.

To start the annotation, we simply connect a laptop with a TV and start the script and a movie via VLC-player. The annotators can also stop and continue the movie as they wish without deranging the synchronization. The final output is currently a simple table with the value of the slider for every 100ms.

3 Annotation Study

To validate and compare the live sentiment annotation approach we conducted an annotation study comparing live annotation with textual annotation. Five different movies were annotated by three different annotators separately from each other for each method leading to 30 annotations, 15 for each method. We then compare the performance in sense of time needed and annotation metrics as well as the subjective experience of the annotators measured via a questionnaire and an interview. The entire study was performed in German and only German material was used since all annotators' first language is German.

3.1 Sample

Ten annotators (7 female, 3 male) participated in the study and we split the annotation in a way that every annotator annotated at most three different movies and at least one textual and one live annotation so every annotator experienced the difference. The order of the movies and the annotation types was counterbalanced to compensate for learning effects and no annotator annotated the same movie twice. The age of the annotators ranged from 25 to 31 ($M= 26.5$). All annotators were students in Digital

Humanities or similar education programs. Participation was voluntary. We interviewed for knowledge of the movies and divided the annotation in a way that every annotator had either no knowledge of the content of the movie or very minor knowledge.

3.2 Material

We selected five different movies from varying genres and epochs to avoid the possibility of specific annotation problems due to these factors: *Rear Window* (1954; thriller), *Christmas Vacation* (1989; comedy), *Scream* (1996, horror), *Avengers* (2012, action), *The Fault in Our Stars* (2014; Drama). We decided to use commercial Hollywood movies since they are subject of our own research. We acquired the DVDs of these movies via our institutional library. We used the German subtitles and transformed them to a simple list of subtitles in a table. Please note that what characters are actually saying and what is displayed by subtitles is not always exactly the same.

3.3 Textual Sentiment Annotation

All annotators that performed textual sentiment annotation received a table with a subtitle per line as well as a summary of the movie and an annotation instruction. In a first meeting the annotators were introduced in the annotation process¹. Annotators had to mark the sentiment expressed by the character that is saying the subtitle on a scale from -5 (very negative) to 0 (neutral) to 5 (positive). We chose this differentiated scale since it resembles the live sentiment annotation more than a nominal annotation. Annotators received this table as xls-file and had one week to complete the annotation but were recommended to perform the annotation in one go. Further, they were instructed to not watch the movies.

3.4 Live Sentiment Annotation

The live sentiment annotation was performed in a media lab at our university. Annotators were sitting on a couch in front of a TV using the annotation slider while watching the movie. They were instructed in the process and the functionality and went through a short trial phase for a short movie. Annotators were instructed to mark the expressed sentiment of the characters seen on the screen via the slider on a scale from 0% (very negative) to 100% (very positive). A test coordinator was present but stayed in the background for the entire viewing. It was possible to take breaks by contacting the test coordinator.

¹ The entire study was performed previous to the COVID pandemic, therefore many steps of the study included in-person meetings.

3.5 Post Annotation Questionnaire and Interview

All annotators had to fill out an online questionnaire after each annotation. Next to demographic information we asked for the time needed for the annotation, how difficult the annotation was perceived on a scale from 1 (not difficult at all) to 7 (very difficult) and how certain one was about the annotation on a scale from 1 (very unsure) to 7 (very sure). We further used the NASA Task Load Index [13] (NASA-TLX) to get a value for the perceived cognitive and physical effort. This is an established questionnaire in psychology [14] consisting of 6 questions about the perceived effort resulting to a scale on 6 (very low effort) to 60 (very high effort). We added open-ended questions into the questionnaire in which annotators could give feedback on problems, challenges and the overall perception of the annotation process. Lastly, we performed a short semi-structured interview with the annotators asking about the perceived difficulties and problems.

4 Results

4.1 Time

We inquired about the exact time needed for annotation without breaks. The average time needed for the textual annotation is 123 minutes, so around 2 hours, while the average for the live annotation is 109 minutes (which is basically the average of the length of all films). However, this difference is not significant as shown with a *Mann-Whitney* test for independent samples and a significance level of $p=0.05$ ($U=-1.12$, $p=.235$). We also asked for the time needed with breaks in the questionnaire showing that textual annotators took most of the time multiple breaks while the maximum break a live annotator took was one short break.

4.2 Annotation Metrics

Agreement among annotators is an important factor in annotation. High agreement is beneficial for later machine learning approaches and also validates the theoretical idea of the annotation. We investigated if the agreement and therefore the overall understanding of the sentiment annotation changes by annotation modality by looking (1) at *Fleiss' Kappa* [11], an established agreement metric for annotations by more than two annotators and (2) the percentage of agreements among annotators.

We transform all annotations into the three classes negative, neutral and positive which is common in sentiment annotation. For the textual annotation we regard -5 to -1 as negative, 0 as neutral and 1-5 as positive and we regard every subtitle as one data point. To keep the data points comparable between the live annotation and the textual annotation we made use of the following heuristic: We regarded the exact time frames in which a subtitle was spoken as data point for the analysis. For this time frame we calculated the average of all annotations we received (since we measure the live annotation in 100ms intervals). We then regard an average of 0-40% as negative 41-59%

as neutral and 60-100% as positive. Annotators reported that it is difficult to mark exactly 50% for neutral so we increased the neutral area. Please note however that by this heuristic we do neglect any annotations that are done outside of the time frames of subtitles. This step is necessary since the agreement statistic reacts sensitive on varying numbers of data points. Table 1 shows the agreement metrics:

Table 1. Agreement metrics per movie and overall.

Movie	Fleiss' Kappa (Text)	Fleiss' Kappa (Live)	Percentage (Text)	Percentage (Live)
The Fault in Our Stars	0.26	0.42	52.15	61.44
Christmas Vacation	0.41	0.35	63.66	51.95
Scream	0.34	0.40	60.83	64.34
The Avengers	0.14	0.10	45.35	48.56
Rear Window	0.06	0.33	44.11	57.91
Average	0.29	0.32	53.22	56.84

The results show that the agreement is slight (0.0-0.2) to fair (0.21-0.40) according to [24], which is rather low but very common in sentiment annotation of narrative and artistic art due to the subjective nature of the task [1, 41, 42, 49]. While there are strong differences for some movies, the averages are just slightly increased for the live sentiment annotation.

4.3 Post Annotation Questionnaire

Table 2 illustrates the results for the perceived difficulty and security (scale from 1-7) and the perceived effort operationalized via the NASA-TLX (6-60).

Table 2. Post Annotation Questionnaire results.

Annotation Type	Perceived Difficulty		Perceived Certainty		Perceived Effort (NASA-TLX)	
	<i>Avg</i>	<i>Std</i>	<i>Avg</i>	<i>Std</i>	<i>Avg</i>	<i>Std</i>
Textual	5.1	1.55	2.6	1.29	35.7	7.41
Live	2.53	1.3	4.2	1.57	31.6	6.86

A Mann-Whiney test of significance for independent samples for all three variables shows that annotators perceived the textual annotation as significantly more difficult ($U=-3.6, p<.001$) and are less certain when annotating textual ($U=-2.6, p=.008$). While the NASA-TLX shows to be of average perceived effort for both types a Mann-Whitney-U Test also shows that the difference is significant ($U=-2.8, p=.004$).

4.4 Post Annotation Interview

The qualitative analysis of the open-ended questions and the interviews led to multiple insights. Participants validated the low agreement metrics by describing the task as challenging and open to interpretation no matter the annotation type. For the textu-

al annotation, participants explicitly criticized the lack of the video channel and the resulting missing context. Problems for the live annotation included how to interpret differing sentiments from different characters on the screen and how to react on fast changes. The textual annotation was described as “very boring” and “exhausting”. While the feedback for the live annotation was not that negative, participants did note that the annotation needs “a lot of concentration throughout” the view. When instructed to compare both approaches all annotators preferred the live sentiment annotation.

5 Discussion

The results of our annotation study are mixed concerning the advantages of live sentiment annotation. Regarding the agreement statistics we did not identify a remarkable difference. The agreement of annotators remains very small, showing again the difficulty and inherent subjectivity of sentiment annotation [1, 41, 42, 49]. While problems of textual sentiment annotation are solved (e.g. the missing visual context) new problems arise like how to deal with fast changes of scenery and how to react to multiple characters with different emotions. One limitation might also be that we adjusted the agreement analysis of the live annotation to the subtitles in the sense that we took the presentation time of the subtitles as units for the analysis neglecting passages without subtitles. Furthermore, these time units are sometimes quite short which might cause problems for the live annotation.

Nevertheless, Kajava et al. [19] were able to achieve rather high annotator agreements in an emotion annotation tasks on subtitles via gamification and the deliberate removal of context showing that the solutions of the agreement problems might point towards gamification and simplification (at least for text). It is also worth noting that the movies annotated are quite long (around 2 hours); thus, requiring a lot of concentration and are more prone to errors than shorter films might be. We plan to examine other annotation types and shorter material to see if we can find differences in the results.

Nevertheless, we could find significant differences in the perceived difficulty, certainty and effort for the annotation task. The live annotation was perceived as more enjoyable than the textual annotation. Feedback of qualitative analysis validated this finding. While both annotation types are not experienced as fun, the live sentiment annotation was preferred by all annotators due to the limited time necessary and the less exhausting experience. However, the task was still experienced as “work” necessary of constant concentration. We still feel encouraged to continue our work to investigate the possibilities of live annotation since the annotation process was overall perceived positive. Agreement statistics are strongly dependent on the validity of the theoretical concept to annotate, the training of the annotators and the clarity of the annotation instruction. Thus, we want to investigate how long-term studies can show an improvement concerning agreements. Please also note that our study is rather small scale in means of number of movies and annotators. Due to legal reasons the possibilities of scaling the study up by performing a similar annotation process online are limited, thus we will focus on public domain movies for our next investigations. An-

other question is if there are still possibilities to reduce the workload of annotators even more and perform the “annotation” fully intuitive by using physiological metrics of the movie viewer (e.g. via skin sensors or facial recognition). For example, in other settings researcher use facial and voice emotion recognition to predict metrics [12, 47]. Using physiological metrics would be a way to bypass problems concerning interpretation biases.

In summary, we come to the conclusion that the selection of the most beneficial annotation type is dependent of the research goal. If one is solely interested in the analysis of the spoken word and context-free sentences ([19]), the inclusion of the video channel might not be helpful and can even be disturbing. However, in our project, we are explicitly interested in the sentiment expressed by characters which is not always easy to identify solely based on the text. We also want to highlight that textual analysis neglects any scenes that do not include spoken words, which can consist of very long time spans. Furthermore, the application of computational methods influences the decision for an annotation approach as well. Textual annotation might certainly be sufficient for solely textual machine learning approaches but the exploration of multimodal approaches is dependent of multimodal annotations to keep the concept concise. While multimodal annotation tools exist, we argue that our live annotation approach delivers benefits for the experience of annotators and we thus plan to continue our research.

References

1. Alm, C.O., Roth, D., Sproat, R.: Emotions from text: machine learning for textbased emotion prediction. In: Proceedings of the conference on human language technology and empirical methods in natural language processing, pp. 579–586. Association for Computational Linguistics (2005).
2. Alm, C.O., Sproat, R.: Emotional sequencing and development in fairy tales. In: International Conference on Affective Computing and Intelligent Interaction, pp. 668–674. Springer (2005).
3. Benikova, D., Biemann, C., & Reznicek, M: NoSta-D Named Entity Annotation for German: Guidelines and Dataset. In: LREC, pp. 2524-2531. (2014).
4. Bird, F. & Liberman, M.: A formal framework for linguistic annotation. *Speech communication*, 33(1-2), 23–60 (2001).
5. Bosco, C., Allisio, L., Mussa, V., Patti, V., Ruffo, G.F., Sanguinetti, M., Sulis, E.: Detecting happiness in italian tweets: Towards an evaluation dataset for sentiment analysis in felicità. In: 5th International Workshop on EMOTION, SOCIAL SIGNALS, SENTIMENT & LINKED OPEN DATA, ES3LOD 2014, pp. 56–63. European Language Resources Association (2014).
6. Bornstein, A., Cattani, A., & Dagan, I.: CoRefi: A Crowd Sourcing Suite for Coreference Annotation. arXiv preprint arXiv:2010.02588. (2020).
7. Burghardt, M.: Usability recommendations for annotation tools. In: *Proceedings of the Sixth Linguistic Annotation Workshop*, pp. 104-112. Association for Computational Linguistics (2012).
8. Burghardt, M., Heftberger, A., Pause, J., Walkowski, N. O., & Zeppelzauer, M.: Film and Video Analysis in the Digital Humanities—An Interdisciplinary Dialog. *Digital Humanities Quarterly*, 14(4), (2020).

9. Chu, E., Roy, D.: Audio-visual sentiment analysis for learning emotional arcs in movies. In: 2017 IEEE International Conference on Data Mining (ICDM), pp. 829–834. IEEE (2017).
10. Dutta, A., Zisserman, A.: The via annotation software for images, audio and video. In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 2276–2279 (2019).
11. Fleiss, J. L., Levin, B., & Paik, M. C: Statistical methods for rates and proportions. John Wiley & Sons (2013).
12. Halbhuber, D., Fehle, J., Kalus, A., Seitz, K., Kocur, M., Schmidt, T. & Wolff, C.: The Mood Game - How to use the player's affective state in a shoot'em up avoiding frustration and boredom. In: Alt, F., Bulling, A. & Döring, T. (Hrsg.), *Mensch und Computer 2019 - Tagungsband*. New York, ACM (2019).
13. Hart, S. G., & Staveland, L. E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: *Advances in psychology*, Vol. 52, pp. 139-183. North-Holland (1988).
14. Hart, S. G.: NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the human factors and ergonomics society annual meeting, Vol. 50, No. 9, pp. 904-908. Sage CA, Los Angeles, CA, Sage publications (2006).
15. Heftberger, A. (2018). *Digital Humanities and Film Studies: Visualising Dziga Vertov's Work*. Springer International Publishing (2018).
16. Hielscher, E.: The Phenomenon of Interwar City Symphonies: A Combined Methodology of Digital Tools and Traditional Film Analysis Methods to Study Visual Motifs and Structural Patterns of Experimental-Documentary City Films. *DHQ: Digital Humanities Quarterly*, 14(4), (2020).
17. Hoff, K., & Preminger, M.: Usability testing of an annotation tool in a cultural heritage context. In *Research Conference on Metadata and Semantics Research*, pp. 237-248. Springer, Cham (2015).
18. Kahneman, D.: *Thinking, fast and slow*. Macmillan (2011).
19. Kajava, K., Öhman, E., Piao, H., & Tiedemann, J.: Emotion Preservation in Translation: Evaluating Datasets for Annotation Projection. In: *DHN*, pp. 38-50, (2020).
20. Kakkonen, T., Kakkonen, G.G.: Sentiprofiler: Creating comparable visual profiles of sentimental content in texts. In: *Proceedings of the Workshop on Language Technologies for Digital Humanities and Cultural Heritage*, pp. 62–69, (2011)
21. Kim, E., & Klinger, R.: An Analysis of Emotion Communication Channels in Fan Fiction: Towards Emotional Storytelling. arXiv preprint arXiv:1906.02402. (2019).
22. Kim, E., Klinger, R.: A survey on sentiment and emotion analysis for computational literary studies. arXiv preprint arXiv:1808.03137. (2018).
23. Kim, E., Padó, S., Klinger, R.: Prototypical emotion developments in literary genres. In: *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pp. 17–26, (2017).
24. Landis, J. R., & Koch, G. G.: The measurement of observer agreement for categorical data. *biometrics*, 159-174 (1977).
25. Liu, B.: *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press (2016).
26. Martin, J.C., Kipp, M.: Annotating and measuring multimodal behaviour-tycoon metrics in the anvil tool. In: *LREC. Citeseer* (2002).
27. Mohammad, S.: From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In: *Proceedings of the 5th ACL-HLT Workshop on Language*

- Technology for Cultural Heritage, Social Sciences, and Humanities, pp. 105–114. Association for Computational Linguistics (2011)
28. Mohammad, S. M.: Challenges in sentiment analysis. In: *A practical guide to sentiment analysis*, pp. 61-83. Springer, Cham (2017).
 29. Momtazi, S.: Fine-grained german sentiment analysis on social media. In: *LREC*, pp. 1215–1220. Citeseer (2012)
 30. Nalisnick, E.T., Baird, H.S.: Character-to-character sentiment analysis in shakespeare's plays. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Volume 2: Short Papers*, pp. 479–483. (2013)
 31. Neves, M., & Ševa, J.: An extensive review of tools for manual annotation of documents. *Briefings in Bioinformatics*. (2019)
 32. Öhman, E.: Challenges in Annotation: Annotator Experiences from a Crowdsourced Emotion Annotation Task. In: *DHN*, pp. 293-301. (2020).
 33. Öhman, E., Kajava, K.: Sentimentator: Gamifying fine-grained sentiment annotation. In: *DHN*, pp. 98–110 (2018).
 34. Öhman, E., Kajava, K., Tiedemann, J., Honkela, T.: Creating a dataset for multilingual fine-grained emotion-detection using gamification-based annotation. In: *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 24–30. (2018).
 35. Pianzola, F., Reborá, S., & Lauer, G.: Wattpad as a resource for literary studies. Quantitative and qualitative examples of the importance of digital social reading and readers' comments in the margins. *PloS one*, 15(1), e0226708. (2020).
 36. Reagan, A.J., Mitchell, L., Kiley, D., Danforth, C.M., Dodds, P.S.: The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(1). (2016).
 37. Salmi, H., Laine, K., Röpötti, T., Kallioniemi, N., & Karvo, E.: Crowdsourcing Metadata for Audiovisual Cultural Heritage: Finnish Full-Length Films, 1946-1985. In: *DHN*, pp. 325-332. (2020).
 38. Schmidt, T.: Distant Reading Sentiments and Emotions in Historic German Plays. In: *Abstract Booklet, DH_Budapest_2019*, pp. 57-60. Budapest, Hungary (2019).
 39. Schmidt, T., Burghardt, M.: An evaluation of lexicon-based sentiment analysis techniques for the plays of Gotthold Ephraim Lessing. In: *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pp. 139–149. Association for Computational Linguistics (2018).
 40. Schmidt, T. & Burghardt, M.: Toward a Tool for Sentiment Analysis for German Historic Plays. In: Piotrowski, M. (ed.), *COMHUM 2018: Book of Abstracts for the Workshop on Computational Methods in the Humanities 2018*, pp. 46-48. Lausanne, Switzerland: Laboratoire lausannois d'informatique et statistique textuelle (2018).
 41. Schmidt, T., Burghardt, M., Dennerlein, K.: Sentiment annotation of historic german plays: An empirical study on annotation behavior. In: Kübler, S., Zinsmeister, H. (eds.) *Proceedings of the Workshop for Annotation in Digital Humanities (annDH)*, pp. 47–52. Sofia, Bulgaria (2018).
 42. Schmidt, T., Burghardt, M., Dennerlein, K. & Wolff, C.: Sentiment Annotation in Lessing's Plays: Towards a Language Resource for Sentiment Analysis on German Literary Texts. In: *2nd Conference on Language, Data and Knowledge (LDK 2019). LDK Posters*. Leipzig, Germany (2019).
 43. Schmidt, T., Burghardt, M. & Wolff, C.: Towards Multimodal Sentiment Analysis of Historic Plays: A Case Study with Text and Audio for Lessing's Emilia Galotti. In:

- Proceedings of the DHN (DH in the Nordic Countries) Conference, pp. 405-414. Copenhagen, Denmark (2019).
44. Schmidt, T., Hartl, P., Ramsauer, D., Fischer, T., Hilzenthaler, A. & Wolff, C.: Acquisition and Analysis of a Meme Corpus to Investigate Web Culture. In: Digital Humanities Conference 2020 (DH 2020). Virtual Conference (2020).
 45. Schmidt, T., Jakob, M. & Wolff, C.: Annotator-Centered Design: Towards a Tool for Sentiment and Emotion Annotation. In: Draude, C., Lange, M. & Sick, B. (Eds.), INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge), pp. 77-85. Bonn: Gesellschaft für Informatik e.V. (2019).
 46. Schmidt, T., Kaindl, F. & Wolff, C.: Distant Reading of Religious Online Communities: A Case Study for Three Religious Forums on Reddit. In: Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN 2020). Riga, Latvia (2020).
 47. Schmidt, T., Schlindwein, M., Lichtner, K., & Wolff, C.: Investigating the Relationship Between Emotion Recognition Software and Usability Metrics. *i-com*, 19(2), 139-151 (2020).
 48. Schmidt, T., Winterl, B., Maul, M., Scharck, A., Vlad, A. & Wolff, C.: Inter-Rater Agreement and Usability: A Comparative Evaluation of Annotation Tools for Sentiment Annotation. In: Draude, C., Lange, M. & Sick, B. (Eds.), INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge), pp. 121-133. Bonn: Gesellschaft für Informatik e.V. (2019).
 49. Sprugnoli, R., Tonelli, S., Marchetti, A., Moretti, G.: Towards sentiment analysis for historical texts. *Digital Scholarship in the Humanities* 31(4), 762–772 (2016)
 50. Takala, P., Malo, P., Sinha, A., Ahlgren, O.: Gold-standard for topic-specific sentiment analysis of economic texts. In: *LREC*. vol. 2014, pp. 2152–2157. (2014)
 51. Tsivian, Y.: Cinemetrics, part of the humanities' cyberinfrastructure. (2009). Retrieved from <https://www.degruyter.com/document/doi/10.14361/9783839410233-007/html>
 52. Yavuz, M. C.: Analyses of Character Emotions in Dramatic Works by Using EmoLex Unigrams. In: Proceedings of the Seventh Italian Conference on Computational Linguistics , CLiC-it'20. (2020).