

# What Can the Sentiment of a Software Requirements Specification Document Tell Us?

Colin Werner  
University of Victoria  
Victoria, BC, Canada  
colinwerner@uvic.ca

Ze Shi Li  
University of Victoria  
Victoria, BC, Canada  
lize@uvic.ca

Neil Ernst  
University of Victoria  
Victoria, BC, Canada  
nernst@uvic.ca

**Abstract**—Sentiment analysis tools are becoming increasingly more prevalent in the software engineering research community. In this data showcase paper, we present a set of twenty-two software requirements specification (SRS) documents that have been preprocessed and subsequently analyzed using the Senti4SD sentiment analysis tool. As part of our preliminary research, we compared the result of the sentiment analysis of the SRS documents and other non-related documents and found that the SRS documents were 6% more neutral than other non-related documents. Finally, we also present a number of research questions that we believe the research community might be able to answer using our published data.

## I. INTRODUCTION

Sentiment analysis is a growing field of software engineering research and there is a keen interest in applying sentiment analysis to the requirements engineering domain. Sentiment analysis has been used to analyze the sentiment of support tickets [1], in conjunction with machine learning techniques to predict support ticket escalations [2], assess user’s sentiment towards application features through application reviews [3], and detect a developer’s sentiment, in particular the direction of anger in collaborative software development [4].

Furthermore, at the inaugural International Workshop on Affective Computing for Requirements Engineering [5], much of the discussion unsurprisingly revolved around how affective computing can aide in increasing the efficacy of requirements engineers. In particular, one such discussion considered how sentiment analysis tools may be used in conjunction with a software requirements specification (SRS) document. In fact, it was unknown whether anyone had ever directly analyzed the sentiment of an SRS document. Many of the discussion participants suspected that the sentiment would be overtly neutral; however, this was merely conjecture. Nonetheless, the notion of evaluating SRS documents was an endeavour that we had decided to explore.

## II. DATA PREPROCESSING

To analyze the sentiment of an SRS document, we first needed to obtain at least one SRS document. Fortunately we were able to collect twenty-two SRS documents from various sources, each of which shall remain anonymous for the purposes of identity protection. Each of the twenty-two SRS documents were written in the English language and received

in the PDF format. The preprocessing stage consisted of three steps: 1.) conversion to plain-text, 2.) conversion to UTF-8, and 3.) tokenization into sentences.

The first preprocessing step of converting all the PDF documents to plain-text was completed using the Automator application, which is a built-in default application as part of MacOSX. It is important to note that typical SRS documents often contain a number of various figures and diagrams, such as UML diagrams, that cannot be easily translated to textual notation and thus cannot be analyzed by sentiment analysis tools. The second preprocessing step of converting to UTF-8, if applicable,<sup>1</sup> was completed using the Unix utility `iconv`. The third preprocessing step utilized the `sent_tokenize` function from the Natural Language ToolKit to tokenize the plain-text documents sentence by sentence, the output of which resulted in each line containing a maximum of one sentence. This sentence tokenization was performed as sentiment analysis is best performed on the minimum number of words (i.e. sentences as opposed to paragraphs).

## III. SENTIMENT ANALYSIS

After the preprocessing was completed, the SRS documents were ready for Sentiment Analysis. Due to the technical nature of SRS documents, we opted to use the Senti4SD [6] sentiment analysis tool, which is specifically trained to handle technical terms, whereby many others are not suitable for technical jargon [7]. Alternatively, as a side project it would be an interesting task to compare the results of analyzing the sentiment of these twenty-two SRS documents with Senti4SD and other common sentiment analysis tools, such as SentiStrength; however, this task is left for future work.

Each line, which corresponds to one sentence, of each SRS document was individually analyzed. As a result, each SRS document had a corresponding output sentiment file whereby each line number contained the line number from the SRS document and the predicted sentiment, in particular one of neutral, positive, or negative.

<sup>1</sup>Not all documents required conversion to UTF-8.

#### IV. DATA HIGHLIGHTS

The resulting sentiment analysis data have been published for academic use via GitHub<sup>2</sup>. We are currently working to publish the textual data representing the twenty-two SRS documents; however, this paper was published prior to receiving permission. As an alternative, we encourage and shall willingly accept any reasonable requests and queries pertaining to the SRS textual data that will not reveal any confidential data (e.g. number of words per line).

In order to provide an initial comparative, meaningful analysis of the sentiment of the SRS documents, we performed subsequent sentiment analysis on sixty-nine of the accepted technical papers from the 2018 IEEE 26th International Requirements Engineering Conference (RE'18) [8]. In particular, the exact same preprocessing and sentiment analysis steps were performed and the results of which have been published (the textual data will not be published, but can be obtained with an appropriate subscription).

Overall, both the SRS documents and the RE'18 papers contained an overly neutral context, whereby the average percentage of neutral lines was 96% and 90% respectively. Interestingly, the SRS documents are on average 6% more neutral than the RE'18 documents. The data for both sets of documents is summarized in Table I.

TABLE I  
SENTIMENT ANALYSIS DATA SUMMARY

		SRS	RE'18
Neutral	Average	96%	90%
	Median	96%	91%
	Minimum	92%	62%
	Maximum	99%	100%
Positive	Average	3%	8%
	Median	3%	6%
	Minimum	1%	0%
	Maximum	6%	32%
Negative	Average	1%	2%
	Median	1%	2%
	Minimum	0%	0%
	Maximum	3%	13%

#### V. IMPLICATIONS AND FUTURE WORK

The implications that can be drawn based on our data are not yet well defined. Part of the incentive for this paper is to publish the sentiment data to enable the community and readers to provide meaningful contributions on how this data can be used. In addition, based on the sample twenty-two SRS documents, we can definitively infer that an SRS has an overly neutral sentiment; however, what this means to the research community is something that needs to be further explored.

Some potential research questions are posed to the audience for future consideration:

- Given the overly neutral sentiment of an SRS document, can sentiment analysis be used as an indicator to evaluate the quality of an SRS document?

- Can sentiment analysis of an SRS document be used as an input feature to classify whether a particular sentence is part of an SRS document?
- Can sentiment analysis be used to as an input feature to classify whether a sentence is part of a SRS domain type?
- Does the sentiment of a SRS document changed based on the specific theme of requirements (e.g. security or privacy are perhaps overly negative)?
- How does the sentiment of an SRS compare to the sentiment of a user story?

#### VI. CONCLUSIONS

Our data showcase paper presents the results of running Senti4SD on a set of twenty-two SRS documents. We also analyzed the sentiment of the accepted RE'18 papers to provide an initial comparison with the twenty-two SRS documents, which shows a 6% difference in the neutral sentiment between the SRS documents and the RE'18 papers. Finally, we propose a number of potential research questions that our published data may be used in order to answer, which we hope will inspire a fruitful discussion and lead to potential answers to our and others' research questions.

#### REFERENCES

- [1] C. Werner et al. "How Angry are Your Customers? Sentiment Analysis of Support Tickets that Escalate". In: *2018 1st International Workshop on Affective Computing for Requirements Engineering (AffectRE)*. 2018, pp. 1–8. DOI: 10.1109/AffectRE.2018.00006.
- [2] C. Werner, Z. S. Li, and D. Damian. "Can A Machine Learn Through Customer Sentiment? A Cost-Aware Approach To Predict Support Ticket Escalations". In: *IEEE Software* (2019), pp. 1–1. ISSN: 0740-7459. DOI: 10.1109/MS.2019.2923408.
- [3] E. Guzman and W. Maalej. "How do users like this feature? a fine grained sentiment analysis of app reviews". In: *2014 IEEE 22nd international requirements engineering conference (RE)*. IEEE. 2014, pp. 153–162.
- [4] D. Gachechiladze et al. "Anger and its direction in collaborative software development". In: *2017 IEEE/ACM 39th International Conference on Software Engineering: New Ideas and Emerging Technologies Results Track (ICSE-NIER)*. IEEE. 2017, pp. 11–14.
- [5] D. Fucci, N. Novielli, and E. Guzmán. "Welcome Message from the Chairs". In: *2018 1st International Workshop on Affective Computing for Requirements Engineering (AffectRE)*. 2018, pp. 6–7. DOI: 10.1109/AffectRE.2018.00005.
- [6] F. Calefato et al. "Sentiment Polarity Detection for Software Development". In: *Empirical Software Engineering* 23.3 (2018), pp. 1352–1382. ISSN: 1573-7616. DOI: 10.1007/s10664-017-9546-9. URL: <https://doi.org/10.1007/s10664-017-9546-9>.
- [7] R. Jongeling, S. Datta, and A. Serebrenik. "Choosing your weapons: On sentiment analysis tools for software engineering research". In: *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. 2015, pp. 531–535. DOI: 10.1109/ICSM.2015.7332508.
- [8] W. Maalej, D. Amyot, and G. Ruhe. "Welcome Message from the RE18 Chairs". In: *2018 IEEE 26th International Requirements Engineering Conference (RE)*. 2018, pp. 13–16. DOI: 10.1109/RE.2018.00005.

<sup>2</sup><https://github.com/The-SEGAL-Group/SRS-Sentiment>