Sentiment Annotation of Slovene User-Generated Content

Darja Fišer,[†*] Jasmina Smailović,[*] Tomaž Erjavec,[*] Igor Mozetič,[*] Miha Grčar[*]

[†]Faculty of Arts, University of Ljubljana
Aškerčeva cesta 2, SI-1000 Ljubljana, Slovenia
darja.fiser@ff.uni-lj.si
[*]Department of Knowledge Technologies
Jožef Stefan Institute
Jamova cesta 39, SI-1000 Ljubljana
jasmina.smailovic@ijs.si, tomaz.erjavec@ijs.si,
igor.mozetic@ijs.si, miha.grcar@ijs.si

Abstract

The Janes corpus contains posts from five different platforms (tweets, forums, blogs, comments on news articles and on Wikipedia) containing 167 million words of Slovene user-generated content. We have annotated the texts in the corpus with their sentiment, using a SVM-based sentiment classifier trained on a large collection of Slovene tweets. The paper introduces the classifier and its model for Slovene, gives an evaluation of the assigned scores and an analysis of the sentiment scores assigned to the text types of Janes.

1. Introduction

Sentiment analysis (also referred to as opinion mining) is a type of text analysis which detects opinions, sentiments and emotions about different entities. It can be applied in various scenarios, e.g., analysing public opinion about companies and products, voters' comments and debates regarding political parties, or investors' expectations about stocks, as well as analyses of the prevailing sentiment in written communication (Dodds et al., 2015). The first approaches to sentiment analysis emerged at the beginning of the century, and since then, it has gained increasing attention, esp. due to the massive usage of on-line platforms, such as blogs, forums and social networking services, where people regularly express their emotions about various topics (Liu, 2012; Liu, 2015).

We report on applying a pre-trained sentiment labelling system to a corpus of Slovene user-generated content (UCG), e.g. *ma nimam besed. Dost mam teh slinastih farjev,ki glumjo sirote,dnarja pa ko toče . Da ne pomislim na Zvon,Betnavo.. Plačajo naj p...! // I'm lost for words. I've had it with these sleazy loaded guys playing the broke card. I don't even want to think about the Zvon, Betnava.. Them f... should pay up!*. The goal of the paper is threefold: a) to evaluate the performance of the system on a collection of fairly heterogenous Slovene texts, b) to perform an analysis of the sentiment characteristics and distribution across different types of Slovene UCG and, last but not least, c) to add valuable metadata to the texts contained in the corpus.

2. Automatic Sentiment Labeling

In this section we describe the sentiment classifier that was used to automatically label the Janes corpus. In particular, we briefly outline the SVM-based algorithm for training the sentiment classifier, the manually labeled training data, and the data preprocessing steps.

2.1. Sentiment Classification Algorithm

We employ the Support Vector Machine (SVM) algorithm (Vapnik, 1995) to train a sentiment classification model. More precisely, we use the TwoPlaneSVMbin implementation, which is a three-class extension of the basic two-class SVM and is introduced in Mozetič et al. (2016). The three-class extension is needed to categorize texts into three sentiment classes: negative, neutral and positive.

TwoPlaneSVMbin is a combination of two binary SVM models, where one model separates the negative examples from the neutrals-or-positives while the other separates the positives from the neutrals-or-negatives. Therefore, two SVM hyperplanes are constructed. Additionally, the vector space is partitioned into bins (in our experiments the bin width is 0.2 of the SVM margin), and for each bin the information about the label distribution of the training examples is calculated. In the classification phase, a new example is projected into the vector space and a corresponding bin is determined. If the number of training examples in the bin is equal or higher than 5 and the distance from the hyperplanes is less than two margins, the class of the example is assigned based on the majority class in the bin. Otherwise, the class is determined based on the sides and distances from both hyperplanes in which the example resides.

2.2. Training Dataset

We acquired a large collection of Slovenian tweets during a joint project with Gama System[1]. The tweets were collected through the Twitter API, assisted by the PerceptionAnalytics platform[2]. The collected tweets are general (they do not discuss any particular entity) and were posted from January 2014 to February 2015.

The Twitter data was manually labeled by seven annotators. The resulting dataset contains 112,832 tweets labeled as negative (34,164), neutral (48,458) or positive (30,210). The dataset was used to train the sentiment classification model and to assess the classifier's performance by performing 10-fold cross validation. The details on evaluation are provided in Mozetič et al. (2016). It should be noted that this manually labeled dataset is not publicly available:

[1]http://www.gama-system.si
[2]http://www.perceptionanalytics.net

while it has been used to train the sentiment classifier and evaluate it, we cannot use it for further experiments.

2.3. Data Preprocessing

Before the training and the classification phase, the data is prepared by applying Twitter-specific and standard preprocessing techniques.

The Twitter-specific preprocessing includes: replacing URLs, hashtags, happy emoticons, sad emoticons, different combinations of punctuation marks, and mentions of Twitter users with common tokens; appending common tokens, which reflect the tweet length or provide information that a tweet contains a stock symbol or a term in uppercase; removing repetitive letters and appending a common token, which represent that a term contained repetitive letters; and normalizing diacritical characters.

The standard text preprocessing techniques consist of performing tokenization, lemmatization, unigram and bigram construction, removing terms which appear less than 5 times in the dataset, and constructing the normalized Delta TF-IDF (Martineau and Finin, 2009) feature vectors.

3. The Sentiment of Janes

The Janes corpus (Erjavec et al., 2015) is the first large (215 million tokens) corpus of Slovene user-generated content. While the corpus is still under construction, the current version of the corpus, Janes v0.4, already contains almost all the texts of the planned final version. The corpus is composed of the following text types:

- **BLOGp**: Blog posts from two popular platforms in Slovenia (*www.rtvslo.si* and *www.publishwall.si*);
- **BLOGc**: Comments on posts in BLOGc;
- **FORUM**: Posts on three popular Slovenian forums (*www.avtomobilizem.com* discussing cars, *med.over.net* on medical and related questions, and *forum.kvarkadabra.net* on scientific topics);
- **NEWS**: Comments on news articles in three popular Slovenian news sites (*www.rtvslo.si*, the portal of the national TV and radio, *www.mladina.si*, the main left-wing weekly magazine, and *www.reporter.si*, the main right-wing weekly magazine);
- **TWEET**: Tweets of 8,749 Slovene users in the period July 2013 – December 2015;
- **WIKI**: Pagetalk and usertalk pages from the Slovene Wikipedia.

Each text in the corpus is richly anntotated with metadata (e.g. author, title, time of post and, of course, sentiment score). Its content has also been linguistically annotated with a tool-chain that consists of rediacritisation, word-form normalisation, part-of-speech tagging and lemmatisation.

The texts in the Janes corpus were automatically annotated also for sentiment, using the SVM model as described in Section 2. As noted, the SVM training set is unavailable and is distinct from the Janes TWEET corpus, and we were interested how the system performs on our data, be it Tweets or other text types contained in Janes.

3.1. Sentiment by Text Type

In order to gain insight into the sentiment-annotated corpus, we created 18 subcorpora with texts of negative, positive, or neutral sentiment for each text type. As Table 1 shows, the largest of these subcorpora is the corpus of tweets with neutral sentiment. At the other end of the spectrum is the almost thirty times smaller subcorpus of wiki posts with positive sentiment. In all text types apart from tweets and wiki posts, negative content dominates. The smallest amount of positive as well as neutral content is in blog and news comments. Positive content prevails only in wiki posts while tweets are predominantly neutral.

Subcorpus	Senti	Tokens	%
BLOGp	neg	12,758,383	**72**
	neut	3,172,827	18
	pos	1,889,522	11
	total	17,820,732	100
BLOGc	neg	11,071,184	**69**
	neut	2,602,217	16
	pos	2,335,223	15
	total	16,008,624	100
FORUM	neg	25,529,662	**55**
	neut	12,715,683	27
	pos	8,053,284	17
	total	46,298,629	100
NEWS	neg	10,765,972	**74**
	neut	2,295,678	16
	pos	1,570,667	11
	total	14,632,317	100
TWEET	neg	32,493,298	34
	neut	36,092,424	**38**
	pos	26,202,339	28
	total	94,788,061	100
WIKI	neg	1,304,319	17
	neut	1,745,448	23
	pos	4,536,936	**60**
	total	7,586,703	100

Table 1: Sizes of the created subcorpora.

These results reflect the differences in the communicative role and nature of the various social platforms. While bloggers and commentators mostly use these on-line channels to express their opinions, disagreement and frustration with the daily politics and other events, forum members and Twitter users focus more on sharing information, news and knowledge, and Wikipedia editors prioritise community building efforts with supportive, encouraging and inclusive communication.

3.2. Sentiment by Key Words

The top key words reflect the domain of the focus corpus very well and can be used to explore differences between corpora (Kilgarriff, 2012). This is why we performed an analysis of 100 top-ranking key lemmas wrt. the complete corpus of that text type. The keyness score of a word is calculated according to the following formula:

$$\frac{fpm_f + n}{fpm_r + n}$$

where fpm_f is the normalized (per million) frequency of the word in the focus corpus, fpm_r is the normalized (per million) frequency of the word in the reference corpus, and n is a smoothing constant, with $n = 1$ the default value.

The key lemmas were manually classified — not taking into account the context they appear in — as positive, negative or neutral. Since they can be used either positively or negatively, proper names, place names and usernames were annotated as neutral lexical items. Mistokenised or mislemmatised words for which it was not possible to determine what they refer to out of context as well as noise in the form of URLs and foreign words were assigned an ”other” tag. Intuitively, most keywords from a subcorpus of texts with negative sentiment would be expected to be negative, etc.

The confusion matrix presented in Table 2 show that subcorpora of tweets best follow this premise as the predominant category of key words is of appropriate sentiment in each subcorpus. The results are very good for all subcorpora of news and blog comments as well. Negative and neutral forum posts behave very well too while top-ranking keywords in the subcorpus of positive comments contain a little more out-of-context neutral words than positive ones. The positive supcorpus of wiki posts is slightly biased towards neutral key words while blog posts display the heaviest bias towards neutral expressions in both the negative and the positive subcorpus, suggesting our automatic sentiment analysis to be the least reliable for this text type.

		L_{neg}	L_{neut}	L_{pos}	Other
BLOGp	neg	21	**77**	1	1
	neut	4	**92**	4	0
	pos	0	**94**	6	0
BLOGc	neg	**64**	32	0	4
	neut	7	**77**	12	4
	pos	0	40	**57**	3
FORUM	neg	**98**	1	0	1
	neut	0	**97**	0	3
	pos	0	**59**	39	2
NEWS	neg	**92**	8	0	0
	neut	1	**75**	8	16
	pos	0	43	**57**	0
TWEET	neg	**99**	1	0	0
	neut	2	**89**	7	2
	pos	0	26	**74**	0
WIKI	neg	**58**	36	4	2
	neut	8	**84**	3	5
	pos	3	**83**	6	8

Table 2: Classification of 100 top-ranking positive, negative and neutral key lemmas for each subcorpus.

The fact that the results for Twitter subcorpora are the best is not surprising, given that the model for sentiment annotation was trained on tweets. News and blog comments which perform second best are not very different from tweets in terms of their length and usage and therefore seem to be almost equally reliably annotated with sentiment. Forum posts and wiki comments are longer but also deal with more specialized topics and have a different communicative purpose and target audience, which is why they are probably harder to annotate with a model trained on twitter data. The biggest outlier in terms of the results but also the most different as a text genre are blog posts.

		N	V	Adj	Adv	F	Np	Oth
BLOGp	neg	**51**	16	16	8	0	5	4
	neut	**38**	0	29	1	0	30	2
	pos	**66**	17	0	17	0	0	0
BLOGc	neg	**33**	**33**	24	10	0	0	0
	neut	32	3	16	1	4	**39**	5
	pos	**27**	4	16	26	**27**	0	0
FORUM	neg	**49**	22	25	4	0	0	0
	neut	**57**	1	31	1	1	7	2
	pos	18	10	**31**	21	20	0	0
NEWS	neg	**45**	27	16	8	3	1	0
	neut	29	1	12	0	3	**40**	15
	pos	**30**	7	18	15	**30**	0	0
TWEET	neg	**37**	21	35	5	2	0	0
	neut	**47**	8	18	6	1	18	2
	pos	30	8	**42**	12	8	0	0
WIKI	neg	**60**	16	19	3	2	0	0
	neut	**51**	0	19	2	0	22	6
	pos	**66**	17	17	0	0	0	0

Table 3: Distribution of keywords per part of speech.

3.3. Sentiment by Part of Speech

For a more detailed understanding of the linguistic nature of the most characteristic vocabulary of positively or negatively charged or neutral texts we performed a part of speech analysis of the 100 top-ranking key lemmas in all the 18 subcorpora. Part of speech assignment was manual. The part of speech was assigned even in cases of lemmatization or tokenization errors. In ambiguous cases, the most common part of speech was assigned. Apart from the main parts of speech, in particular nouns, verbs, adjectives, and adverbs, proper nouns (Np) were considered as a separate category because they were very prominent in certain subcorpora and called for a more detailed treatment. Non-Slovene words were annotated with a ”foreign” (F) tag. Pronouns, conjunctions, abbreviations and interjections were also annotated, but were so infrequent in all subcorpora that they were subsequently merged into a single category called ”other”.

As can be seen in Table 3, nouns, with a 18-66% share, are the most prevalent overall. It is interesting, however, that different parts of speech are most indicative for different sentiments. All negative sentiment subcorpora display the highest proportion of top-ranking key nouns and a much higher proportion of verbs than subcorpora of positive or neutral sentiment. While positive sentiment subcorpora too contain a high proportion of nouns, the most prevalent part of speech in tweet and forum positive sentiment subcorpora are adjectives. Adverbs have the largest share in positively charged news and blog comments while blog comments and forum posts also contain a significant number of adverbs. Proper nouns figure by far the most frequently in neutral sentiment subcorpora, especially in neutral news comments where at 40% they are the most frequent category. Neutral news comments are the only category with a significant share of abbreviations (15%) while foreign words, conjunctions and pronouns were all very rare this high on the key word lists for all subcorpora.

These results suggest that we use different linguistic means for communicating different sentiment. Negatively charged messages will be expressed directly, with nouns and verbs, while positive messages will be delivered descriptively, through adjectives and adverbs. Neutral, factual an informative content is characterized by frequent mentions of persons and their titles.

3.4. Sentiment Lexica

Finally, the 100 top-ranking key lemma lists from all positive and negative sentiment subcorpora were used to build sentiment lexica. Only the key lemmas that were manually annotated as negative or as positive were taken into account. All such lemmas were collected from all 5 subcorpora for each sentiment and added to the lexica. The negative sentiment lexicon created in this way contains 263 different words, 36 (14%) of which appear in three or more subcorpora. 44% of the lemmas in the lexicon are nouns, 25% each are adjectives and verbs, and 6% adverbs. The only two words that appeared in all five negative sentiment subcorpora are the verb *sovražiti* (hate) and the adverb *brezveze* (nonsense).

It is interesting to note that despite the fact that the keywords in subcorpora with positive sentiment showed a much greater variety in terms of their part of speech than their negative counterparts, the positive sentiment lexicon built in the same way contains only half as many words (146) as the negative one. 12% of these appear in at least three corpora, which is similar to the results in the negative sentiment lexicon. Here too the most frequent category are nouns (40%), followed by adjectives (29%) and adverbs (14%). Unlike in the negative sentiment lexicon where there are not found at all, interjections (9%) are an important part of the positive sentiment lexicon while verbs (7%) are barely present. The only word that appears in all five positive sentiment subcorpora is the interjection *bravo* (well done).

Sentiment lexica with lemmas that appear among the 100 top-ranking key lemmas in at least three subcorpora are listed, together with their translation into English, in Table 4 for negative sentiment and in Table 5 for positive sentiment. As can be seen from the tables, a major part of the negative vocabulary expresses personal stance, discontent with the political situation and the governing elites who are seen as corrupt and incompetent as well as the negative emotions authors of these message experience in response to unfavourable political and economic circumstances. The list also contains offensive and discriminatory words that indicate intolerance towards certain social groups. Positive vocabulary, on the other hand, is distinctly interactive and phatic, suggesting that the main communicative function of messages with positive sentiment is relationship and community building with positive feedback, such as praise, congratulations, thanking and good wishes.

Keyword	English	PoS	Subcorpora
baraba	bastard	N	4
bedarija	rubbish	N	4
drek	shit	N	4
lopov	crook	N	4
sranje	crap	N	4
svinjarija	bullshit	N	4
bruhanje	vomit	N	3
cigan	gypsy	N	3
gnoj	bullshit	N	3
kreten	idiot	N	3
kriminalec	criminal	N	3
laž	lie	N	3
sram	shame	N	3
sramota	shame	N	3
butast	stupid	Adj	4
žalosten	sad	Adj	4
beden	pathetic	Adj	3
bolan	sick	Adj	3
glup	stupid	Adj	3
kriv	guilty	Adj	3
nesposoben	incompetent	Adj	3
obupen	terrible	Adj	3
ogaben	disgusting	Adj	3
pokvarjen	corrupt	Adj	3
zmešan	crazy	Adj	3
sovražiti	**hate**	**V**	**5**
groziti	threaten	V	3
jebati	fuck	V	3
krasti	steal	V	3
nakladati	yack	V	3
pljuvati	spit	V	3
pobijati	kill	V	3
smrdeti	smell	V	3
ubiti	kill	V	3
brezveze	**nonsense**	**Adv**	**5**
žalostno	sad	Adv	3

Table 4: Negative sentiment lexica with keywords from at least three subcorpora.

4. Evaluation of the Sentiment Scores

We have performed a manual evaluation of the automatically assigned sentiment scores on a sample of the corpus. The sample contained random 600 texts, 120 from each text type, except form BLOGc, for which we obtained results consistent which news comments in keyword analysis and therefore did not include them in the dataset as we presumed that here too comments on blogs would be very similar to news comments. In addition, we balanced the number of texts from the sources of particular text types, e.g. for NEWS there are 40 texts from each of *www.rtvslo.si*, *www.mladina.si*, and *www.reporter.si*. This was done to arrive at a more diverse sample, as otherwise the much larger sources would swamp the smaller ones, e.g. the number of comments from *www.reporter.si* is only 5% of the those from *www.rtvslo.si*. It should also be noted that the length of an individual text varies widely between the text types. The shortest are tweets, with an average of 12 words per

Keyword	English	PoS	Subcorpora
čestitka	congratulations	N	4
pohvala	praise	N	3
poklon	bow	N	3
carski	great	Adj	3
dobrodošel	welcome	Adj	3
lep	nice	Adj	3
odličen	excellent	Adj	3
super	super	Adj	3
odlično	excellent	Adv	3
pohvalno	deserving compliment	Adv	3
srečno	good luck	Adv	3
bravo	**well done**	**Adv**	**5**
hvala	thank you	Adv	4
tooo	yesss	Adv	3
tnx	tnx	Adv	3
čestitati	congratulate	V	4
polepšati	make (sbd's day)	V	4

Table 5: Positive sentiment lexica with keywords from at least three subcorpora.

text, followed by news comments (42 words), Wikipedia (51), with blogs being the longest (71).

Each text was manually assigned a sentiment score by three annotators, where the annotators also had the option of marking individual texts as out of scope, as they were in a foreign language or contained e.g. adverts and were thus not user-generated, resulting in the final evaluation sample of 555 texts.

The manually assigned scores were compared to each other while the automatically assigned ones were compared to the majority vote (i.e. the label assigned by the most annotators). The agreement results in terms of Krippendorff's Alpha (Krippendorff, 2012) are given in Table 6. Perfect agreement is reached when $Alpha = 1$, while $Alpha = 0$ indicates agreement by chance. Acceptable inter-annotator agreement for this type of task is estimated at $Alpha > 0.4$ (Mozetič et al., 2016).

	All	Wiki	News	Blog	Forum	Tweet
Humans	0.563	0.464	0.513	0.594	0.464	0.547
Auto-major	0.432	0.402	0.394	0.446	0.245	0.372
n	555	107	115	115	119	99

Table 6: The agreement measures in terms of Krippendorff's Alpha for different sub-samples of the corpus which contain n texts.

The table confirms that assigning sentiment scores is a very subjective task and difficult to perform automatically. All the interannotator agreements are below 0.6 Alpha, which, while acceptable, is far from perfect agreement. The automatic assignment of sentiment labels is, of course, worse than the agreement between humans; while it is, overall, above the acceptability threshold, it is slightly below it for three out of five text types. However, it should be noted that the evaluation of the automatic system was quite strict, as it was compared to the majority class of the human annotators, i.e. even in cases the humans did not agree on the score, the system was penalised when it disagreed with the majority vote.

Blogs seem to be the easiest to assign a sentiment to, as both humans and the automatic assignment achieve here the highest score. This is most likely due to the length of the text, where it becomes clear which overall sentiment is expressed by the author. For humans, the second easiest are tweets, whereas the automatic system preforms worse on them than on News and Wiki. This is especially interesting as the automatic system was trained on tweets and would therefore be expected to perform best on the same type of texts. An explanation could be the short length of tweets, which does not give the system enough data to correctly determine the sentiment. Furthermore, it is likely that Twitter is often less straightforwardly opinionated than other types of text, i.e. it contains more ironic posts, which are hard for the automatic system to detect.

In general, the evaluation shows that it might help giving the annotators more precise instructions — preferably in line with those for annotating the training data — with which we would increase the interannotator agreement, while it is less clear on how to improve the quality of the automatic labelling. Here, providing additional training data from the text type that performed the worst, namely Forums, might be of help.

5. Conclusions

The paper presented a sentiment classification system trained on Slovene tweets and its application on the Janes corpus of Slovene user-generated content. The analysis of sentiment-specific keywords gives interesting insight into the vocabulary that is typically used to express different sentiment. Evaluation results show that automatic sentiment classification is consistent with human judgements and that there are considerable differences among the performance of the system across genres. Although the sentiment annotation accuracy could still be significantly improved, the current annotation of the Janes corpus is already useful for e.g., selecting only those texts that have predominantly negative, neutral or positive sentiment and performing on them targeted linguistic analyses.

A detailed analysis of disagreement among the annotators and an error analysis of incorrectly classified texts is planned in the future, which will reveal the outlying problem areas as well as provide clues for further refinements of the algorithm. Another venue of future work is to tackle irony and identify more fine-grained sentiment at the paragraph or even sentence level.

Acknowledgements

The authors thank the two anonymous reviewers for their useful comments and suggestions. We also thank Mojca Mikac for computing Alpha for the manual evaluation, and Sašo Rutar who implemented several classification algorithms and evaluation procedures in the LATINO library for text mining (https://github.com/latinolib). We acknowledge Gama System (http://www.gama-system.si) who collected the tweets for training the sentiment classifier. We thank Sowa Labs (http://www.sowalabs.com) for providing the Goldfinch platform for manual sentiment annotations.

This work was supported in part by the European Union project DOLFINS (no. 640772), by the Slovenian ARRS research project Resources, Tools and Methods for the Research of Nonstandard Internet Slovene (no. J6-6842), and by the ARRS programme Knowledge Technologies (no. P2-103).

6. References

Peter Sheridan Dodds, Eric M. Clark, Suma Desu, Morgan R. Frank, Andrew J. Reagan, Jake Ryland Williams, Lewis Mitchell, Kameron Decker Harris, Isabel M. Kloumann, James P. Bagrow, Karine Megerdoomian, Matthew T. McMahon, Brian F. Tivnan, and Christopher M. Danforth. 2015. Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences*, 112(8):2389–2394.

Tomaž Erjavec, Darja Fišer, and Nikola Ljubešić. 2015. Razvoj korpusa slovenskih spletnih uporabniških vsebin Janes. In *Zbornik konference Slovenščina na spletu in v novih medijih*, pages 20–26, Ljubljana. Znanstvena založba Filozofske fakultete.

Adam Kilgarriff, 2012. *Text, Speech and Dialogue: 15th International Conference, TSD 2012, Brno, Czech Republic, September 3-7, 2012. Proceedings*, chapter Getting to Know Your Corpus, pages 3–15. Springer Berlin Heidelberg, Berlin, Heidelberg.

Klaus Krippendorff. 2012. *Content Analysis, An Introduction to Its Methodology*. Sage Publications, Thousand Oaks, CA, 3rd edition.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.

Bing Liu. 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.

Justin Martineau and Tim Finin. 2009. Delta TFIDF: An improved feature space for sentiment analysis. In *Proc. 3rd AAAI Intl. Conf. on Weblogs and Social Media (ICWSM)*, pages 258–261.

Igor Mozetič, Miha Grčar, and Jasmina Smailović. 2016. Multilingual Twitter sentiment classification: The role of human annotators. *PLoS ONE*, 11(5):e0155036.

Vladimir N. Vapnik. 1995. *The Nature of Statistical Learning Theory*. Springer.