B. Mitschang et al. (Hrsg.): BTW 2017 – Workshopband,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2017 375

Understanding Trending Topics in Twitter

Roland Kahlert Matthias Liebeck Joseph Cornelius

Institute of Computer Science, Heinrich Heine University Düsseldorf, Germany
{roland.kahlert,joseph.cornelius}@hhu.de
liebeck@cs.uni-duesseldorf.de

Abstract: Many events, for instance in sports, political events, and entertainment, happen all over the globe all the time. It is difficult and time consuming to notice all these events, even with the help of different news sites. We use tweets from Twitter to automatically extract information in order to understand hashtags of real-world events. In our paper, we focus on the topic identification of a hashtag, analyze the expressed positive, neutral, and negative sentiments of users, and further investigate the expressed emotions. We crawled English tweets from 24 hashtags and report initial investigation results.

Keywords: Text Mining, Topic Recognition, Sentiment Analysis, Emotion Detection, Twitter

1 Introduction

Social media sites, such as Twitter, Facebook, and Instagram, allow users to share their thoughts, feelings, journals, and travels in the form of text and images. In order to group similar content, these sites provide the functionality to label posts with so-called hashtags. A hashtag consists of a string which is preceded by the character '#', like #fifa. Hashtags can be used for a variety of functions. They may be used to group posts by topics, by emotions (for instance #joy or #love) or by events in the real world (e.g., boxing matches or scandals) which are often discussed by the users.

The growth of social media sites has been enormous over the last years. Since many people use Twitter to discuss events, a system that is able to understand hashtags about events is envisaged. The system could be used on popular hashtags to textually describe what happened in the world, who is concerned about it, and what the public opinion is. It would be timesaving to receive such information as a summary without the need to read and understand hundreds of text messages, which are called "tweets" on Twitter.

There are a few issues that have to be taken into account with such an approach:

(i) Some events may only be discussed in a language other than English. In order for such a system to work, we need language-specific models for natural language processing components and must adapt our techniques to every supported language individually, which requires the ability to understand these languages. We focus on English tweets because of the large amount of available resources for natural language processing.

(ii) Different events can be tagged with the same hashtag. For instance, tweets can be tagged with #WorstDayOfMyLife. Person A might have had a series of bad news on a particular day, whereas person B might have another series of bad things that happened to him or her. As part of their nature, a summarization of such hashtags will fail because they comprise tweets of more than one event.

(iii) Smaller events might not be frequently discussed on social media. For example, the opening of a certain restaurant might be of interest. However, if this is not frequently discussed on Twitter, our system will not process the tweets from the hashtag since we gather tweets from the most popular hashtags.

This paper focuses on our approaches for an automatic understanding of hashtags about events. The remainder of this paper is structured as follows: The next chapter discusses related work about analyzing tweets. In Section 3, we briefly describe our dataset and highlight multiple subtasks which we consider interesting in a hashtag summarization. Thereafter, we report our initial results and exemplarily show good and bad results. We conclude in Section 5 and outline future work.

2 Related Work

The analysis of tweets has gained much popularity in the last years. The topic detection of tweets has been the focus of previous analyses [HD10, OKA10].

Hong and Davison [HD10] used *Latent Dirichlet Allocation (LDA)* [BNJ03] to analyze tweets. The authors focused on two tasks: (i) the prediction of whether a tweet will be retweeted in the future and (ii) the classification of users and their messages into topical categories. For the second task, tweets were crawled from more than 250 verified users who were supposedly posting messages belonging to one of 16 topics, such as *entertainment* or *politics*. Instead of using a fixed set of topics, we focus on automatically describing topics from tweets.

TweetMotif [OKA10] focused on the summarization of tweets that were returned from user queries. The authors used frequent n-grams of length 1 to 3 to extract multiple topic labels that are frequent in the returned tweets but infrequent among other tweets. In order to refine the results, similar topics and near-duplicated tweets were grouped together. TweetMotif's final visualization of a user query consists of multiple topic labels and several exemplary tweets that contain the topic labels.

3 Data and Investigated Tasks

In this section, we describe our Twitter dataset and briefly outline the tasks we focused on for our automated analysis.

3.1 Data

In this paper, we focus on tweets from Twitter that are limited to 140 characters in length. Due to the length restriction, tweets have unique characteristics as users often do not write complete sentences or utilize abbreviations to shorten their text content. Each tweet is posted by a user. We crawled English tweets from 23 different hashtags about events in 2015 and added the non-event hashtag #love. Table 1 lists the crawled hashtags in our dataset and their respective descriptions that we created manually. In total, we crawled roughly 3.3 million tweets.

Hashtag	Tweet count	Timeframe	Event description
#bbking	46275	Jan - Dec	Blues singer B.B. King died
#blatterout	19229	Jan - Dec	People demanding the resignation of Sepp Blatter
#broner	41967	May - Jul	Boxing: Adrian Broner vs Shawn Porter
#charlestonShooting	154061	Apr - Dec	Mass shooting in Charleston, SC
#dieselgate	11344	Feb - Dec	Volkswagen emissions scandal
#endAusterityNow	25875	May - Jul	Protests against austerity in UK
#expo2015	58995	Mar - Dec	Universal Exposition hosted by Milan
#fifa	179310	May - Jul	FIFA corruption scandal
#GameOfThrones	363576	May - Jul	Popular TV series
#germanwings	59684	Jan - Dec	Germanwings flight crashed in the Alps
#heartgate	1057	Jan - Dec	Twitter replaces favorite stars with hearts
#KGL9268	1074	Oct - Dec	Metrojet flight crashed in the Sinai
#love	1043780	May - Jul	All time trending topic
#nbafinals	435817	May - Jul	NBA finals: Warriors vs Cavaliers
#nepalearthquake	121579	Apr - Dec	Heavy earthquake in Nepal
#ohNoHarry	17455	May - Jul	Harry Styles falls off stage
#plutoflyby	32764	Jul	Space probe *New Horizons* reaches Pluto
#PSYAngBatasNgApi	77038	Jun - Dec	1M tweets for #PSYAngBatasNgApi
#seppblatter	22832	Jan - Dec	Sepp Blatter resigns after corruption scandal
#uswnt	261685	Jan - Dec	US women's soccer team wins the world cup
#volkswagen	92438	Jan - Dec	VW, but mainly including #dieselgate
#windows10	241637	Mar - Dec	Release of Windows 10
#wwdc15	52440	Jan - Dec	Apple Developer Conference 2015
#WWEChamber	411	Jun - Dec	World Wrestling: Owens vs Cane

Tab. 1: List of the crawled hashtags in our dataset and descriptions of their content

3.2 Research Tasks

We now briefly describe and motivate the three analysis tasks that we focus on.

3.2.1 Topic Detection

Topic detection is the task of automatically detecting key words or key phrases that describe the topic of text content. In our specific use case, we aim to summarize the discussion topic of each hashtag in our dataset. As our dataset is focused on events, the desired summarization should convey a general understanding of each event. We focus on extractive techniques in order to summarize the content of the hashtag.

The topic detection of a hashtag is challenging. First of all, it is difficult to formally define the term *topic*. In our work, a summarization should contain information regarding an entity that performed an action or is affected by it. For example, let us assume that the following three example sentences are tweets:

> Sebastian Vettel won the Grand Prix.
> Vettel was the winner of the last race.
> The race in Monaco was won by the German driver Sebastian Vettel.

A manual summarization of these tweets could be *"The Monaco Grand Prix was won by the German driver Sebastian Vettel."* The automatically extracted topic should at least comprise a subset of these information. We make the assumption that each hashtag in our dataset, except for #love, contains one topic which should become apparent with our approaches if we have a large enough collection of tweets per hashtag.

However, the topic detection is still challenging because:

(i) Language is versatile since the same content can be described with different words, e.g., with synonyms. Persons or locations can be mentioned with their full names or only with parts of their names, like surnames. Additionally, Twitter users tend to use abbreviations.

(ii) Is it difficult to automatically evaluate the results of an algorithm or to compare the results of two different approaches because this requires a reasonable distance function between the generated output and manually created descriptions.

3.2.2 Sentiment Analysis

Sentiment analysis is the task of identifying positive, negative, and neutral statements in text content. This task is useful in a variety of application domains. For instance, a company might be interested in their customers' opinion on social media sites. Commonly analyzed application domains include customer reviews [HL04] and film reviews [PLV02].

Depending on the individual use case, the sentiments expressed should be analyzed with their viewpoint. As an example, the statement *"The new product of company A is a bad product."* is a negative statement for company A, but it might be a good statement for its competitor.

Sentiment analysis differs between application domains because of domain-specific wordings. For instance, product reviews on Amazon often contain adjectives that describe certain product features, like *"The battery life is really good."* In movie reviews, there are wordings like *"The actor deserves an Oscar for his performance."* which are very domain-specific.

3.2.3 Emotion Detection

In addition to positive and negative sentiments, we want to capture emotions that users express in order to gain a deeper insight into their feelings. Human emotions have been extensively studied in the past. Paul Ekman proposed the existence of basic emotions [Ek92] (anger, disgust, fear, joy, sadness, and surprise) which have been observed in the context of facial expressions. There has been research about other emotion models in the past, such as *Plutchik's wheel of emotions* [Pl80] which comprises the following eight basic emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.

In the case of our dataset, we want to automatically identify emotions in tweets in order to provide a rough overview over the users' feelings, e.g., if they express more sadness than joy.

4 Experimental Results

After describing our dataset and interesting tasks, we now outline our approaches and experimental results.

First, we need to preprocess the tweets in our dataset with a natural language processing pipeline. We use the Twitter-specific pipeline TweeboParser [Ko14] to extract sentences, words, and their part-of-speech tags. Afterwards, we are able to extract task-specific information, such as nouns and verbs for a topic detection.

4.1 Topic Detection

One of the more intuitive solutions to the topic detection of the tweets is to simply count all occurrences of words in a bag-of-words model and report the m most commonly used words in all tweets. Table 2 lists the results of our topic detection approaches for the hashtag #KGL9268. In our case, most of these ten words (listed in the row named *Counter*) are prepositions and articles which do not add any contextual information. They can be regarded as stopwords and be filtered out using a stopword list, which we refer to as *StopWordCount*.

Another idea to express a topic is to only use nouns and names. To accomplish this, we use the part-of-speech tags to filter the words of every tweet. Thereafter, we rank the most frequent words as *NounPOSCount*. We also experiment to describe the storyline by choosing the most frequent verbs (*VerbPOSCount*).

Additionally, we examine if the first sentence of a tweet provides most of the information. We combine our approaches regarding nouns and verbs and only apply them to the first sentence in *FirstSentenceCount*. Our approach *VerbPhraseCount* considers dependencies between the root verbs of each sentence and the nouns in a bag-of-words model.

Approach	Top 10 results
Counter	#kgl9268, the, to, of, in, crash, russian, #egypt, a, plane
StopWordCount	#kgl9268, crash, russian, #egypt, plane, #russia, #7k9268, flight, egypt, sinai
NounPOSCount	egypt sinai, plane, russian spygame, crash, flight, #kgl9268, family, bomb, #egypt, airliner
VerbPOSCount	crash, say, claim, bring, know, cause, break, it', fly, ru
FirstSentenceCount	crash, plane, russian spygame, egypt sinai, flight, #kgl9268, family, #egypt, bomb, condolence
VerbPhraseCount	crash[plane], claim[video], break[plane, apart], ru[poo-poo'd], cause[flight], confirm[offical], rip[soul], mourn[today, people], kill[flight, people], fly[airline]
NGramCount ($n = 3$)	[russian, plane, crash], [sinai, plane, crash], [russian, airliner, crash], [plane, crash, survivor], [survivor, russian, airliner], [crash, russian, flight], [plane, crash, egypt], [survivor, crash, russian], [flight, sinai, egyptian], [russian, flight, sinai]
LDA 1st	plane crash, russian airliner, condolence family, flight crash, thought prayer, crash victim, bbc news, crash survivor, 224 people, family friend
LDA 2nd	russian plane, russian flight, sinai plane, crash russian, crash site, claim responsibility, crash egypt, deep condolence, crash sinai, #kgl9268 crash

Tab. 2: Overview of the topic detection for hashtag #KGL9268

In order to better understand the structure of a tweet, we also extract word sequences of length n from the tweets which were previously filtered to only contain adjectives, numbers, verbs, and nouns (*NGramCount*).

Futhermore, we use *Latent Dirichlet Allocation (LDA)* [BNJ03] to find k topics in a collection of tweets. For each word w from a vocabulary V, LDA calculates the probability $\varphi_{w,t}$ that w belongs to topic t. In our approach, we set $k = 2$, use bigrams, and extract 10 words with the highest probability from each topic. For #KGL9268, the *LDA* approach seems to provide the most information.

We now list several observations that we made during our experiments.

The hashtag #ohNoHarry is a good example for our use case because its name is not self-explanatory. The *NGramCount* approach provides us almost sentence-like n-grams, which tell us that singer Harry Styles fell off a stage and the community was amused of it.

Furthermore, #volkswagen is a good example for multiple topics inside a hashtag. Table 3 shows that LDA is capable of differentiating the 2015 emission scandal from other news regarding the company, e.g., new car models. The results for #love show us that the hashtag does not form a coherent topic. This could be due to the large amount of tweets that do not share the same topic. While the first topic of #nepalearthquake is about asking for help, the second topic deals with the victims.

We also observed for #PSYAngBatasNgApi that Twitter users express pride about one million tweets in the hashtag, but our approaches were unable to tell what exactly the hashtag is about. A reason for this might be the missing contextual information in our collected tweets because fans of a specific entity are not going to describe the entity in their tweets.

Hashtag	First topic	Second topic
#volkswagen	#volkswagen volkswagen, volkswagen beetle, emission scandal, beetle classic, #volkswagen scandal	volkswagen golf, golf gti, vw golf, #volkswagen golf, new #volkswagen
#love	good morning, #love love, #love com, celine dion, #love #quote	#love #photography, #photography #fashion, love #love, love love, #money #love
#nepalearthquake	nepal earthquake, relief effort, need help, #nepalearthquake relief, #nepalearthquake victim	death toll, affect #nepalearthquake, victim #nepalearthquake, thought prayer, people nepal

Tab. 3: Top five *LDA* results for various hashtags

4.2 Sentiment Analysis

Since our corpus does not contain sentiment annotations, we are unable to evaluate the output of any sentiment analysis approach on our dataset. However, SemEval-2016 covered a specific challenge [Na16] for sentiment analysis in Twitter. We decided to use the publically available system from [Gi16] which ranked at fifth place.

[Gi16] classifies each tweet as positive, neutral or negative by using an ensemble of two linear support vector machines (SVM). The first SVM is trained on part-of-speech tags, sentiment lexicons, negations, cluster of tweets, and morphological features. The second SVM uses the centroid of the word embeddings in GloVe [PSM14] of all words in a tweet. The embeddings were pre-trained on tweets.

The sentiment distribution of the hashtags in our dataset is illustrated in Figure 1. The hashtag #expo2015 has a significantly low amount of negative tweets. The most number of negative tweets were posted regarding #charlestonShooting while the amount of neutral tweets remains stable across all hashtags. It is quite surprising to see that more negative sentiments are expressed in the hashtag #heartgate than in #KGL9268 or in #germanwings. This prompted us to further investigate the emotions expressed in these hashtags.

4.3 Emotion Detection

In this work, we use the emotion dictionary *EmoLex* [MT13] which is based on *Plutchik's wheel of emotions* [Pl80] and the following eight basic emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.

Each word in EmoLex is listed as positive or negative and has 8 binary values, which represent whether the word expresses the respective emotion. For example, the word *love* is listed as $[0,0,0,0,1,0,0,0]$. EmoLex was constructed via crowdsourcing on Amazon's Mechanical Turk. In crowdsourcing, small tasks, called *Human Intelligence Tasks* (HIT), are solved by multiple human workers at a very low price.

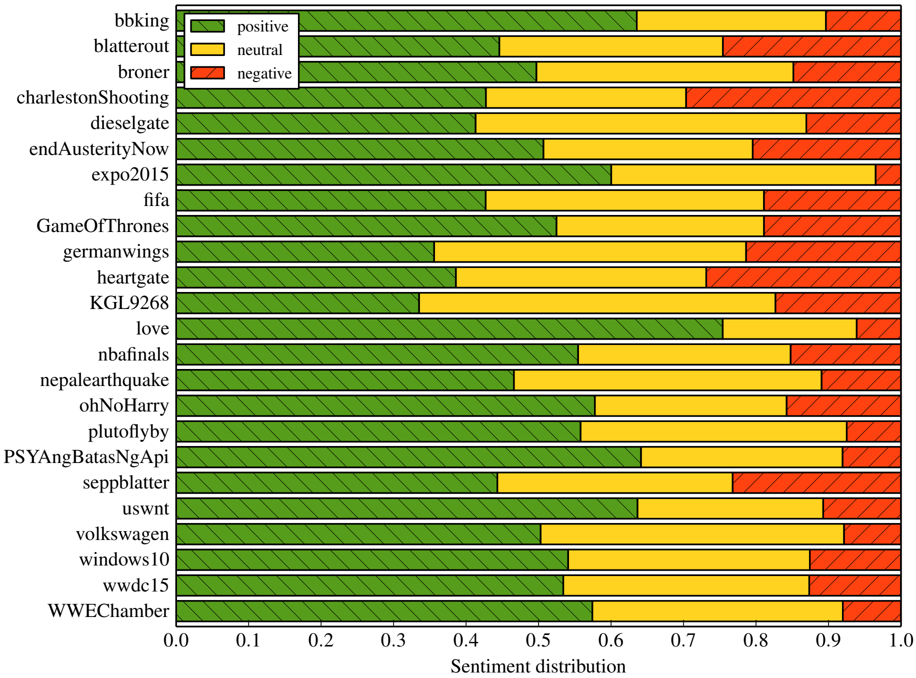

Fig. 1: Summarized sentiments per hashtag

In our analysis of the emotions, we use the lexicon to look up every word from each tweet and calculate an average emotion distribution. Our results are illustrated in Figure 2.

The hashtag with the most emotions expressed is #charlestonShooting where roughly 30% of the tweets are emotional, mostly expressing fear and sadness. The fewest emotions were expressed in the scientific hashtag #plutoflyby, which mostly consists of news-like tweets. It can be observed that fear is often expressed in hashtags that are about events with fatalities, such as airplane crashes (#KGL9268 and #germanwings). Surprisingly, joy is often expressed for #bbking and the tweets in #ohNoHarry contain a high amount of sadness.

5 Conclusion and Future Work

We have presented our approach to an automated understanding of hashtags on Twitter that deal with events. We focused on an extractive topic detection, sentiment analysis, and emotion detection. Regarding topic detection, the more intuitive methods do not provide enough information in order to understand the hashtags. Our *NGramCount* approach and LDA yield short word sequences that help us to get a rough idea of the content. Furthermore, we analyzed expressed sentiments and emotions in the dataset.

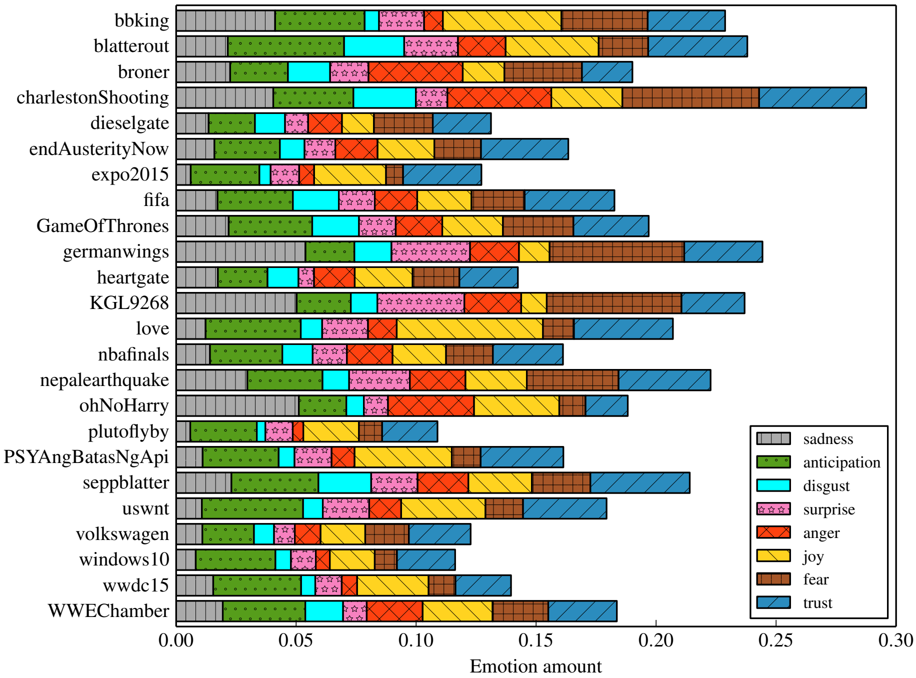

Fig. 2: Summarized emotions per hashtag

In our future work, we want to annotate a subset of our crawled tweets to be able to evaluate techniques for sentiment analysis and to tweak existing approaches for hashtags about events. Furthermore, we will use methods to predict semantic textual similarity between tweets and merge tweets that are almost identical. We aim to utilize the unsupervised *Overlap* method [Li16] and adapt it to Twitter.

Another interesting task is the prediction of user demographics, like age and gender. With such information, a middle-aged man could skip topics that are discussed by younger girls, such as fashion shows. Unfortunately, demographic information about users are not publically available. We would like to use crowdsourcing to annotate demographics for a subset of the users that have posted the tweets. Then, we could incorporate and adapt systems from the *author profiling* challenges in the PAN Series [Ra13], for instance [MLC16], to generate statistics about the users.

References

[BNJ03] Blei, David M.; Ng, Andrew Y.; Jordan, Michael I.: Latent Dirichlet Allocation. Journal of Machine Learning Research, 3:993–1022, 2003.

[Ek92] Ekman, Paul: An Argument for Basic Emotions. Cognition and Emotion, 6(3-4):169–200, 1992.

[Gi16] Giorgis, Stavros; Rousas, Apostolos; Pavlopoulos, John; Malakasiotis, Prodromos; Androutsopoulos, Ion: aueb.twitter.sentiment at SemEval-2016 Task 4: A Weighted Ensemble of SVMs for Twitter Sentiment Analysis. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). Association for Computational Linguistics, pp. 96–99, 2016.

[HD10] Hong, Liangjie; Davison, Brian D.: Empirical Study of Topic Modeling in Twitter. In: Proceedings of the First Workshop on Social Media Analytics. SOMA '10. ACM, pp. 80–88, 2010.

[HL04] Hu, Minqing; Liu, Bing: Mining and Summarizing Customer Reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '04. ACM, pp. 168–177, 2004.

[Ko14] Kong, Lingpeng; Schneider, Nathan; Swayamdipta, Swabha; Bhatia, Archna; Dyer, Chris; Smith, Noah A.: A Dependency Parser for Tweets. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014. pp. 1001–1012, 2014.

[Li16] Liebeck, Matthias; Pollack, Philipp; Modaresi, Pashutan; Conrad, Stefan: HHU at SemEval-2016 Task 1: Multiple Approaches to Measuring Semantic Textual Similarity. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). Association for Computational Linguistics, pp. 607–613, 2016.

[MLC16] Modaresi, Pashutan; Liebeck, Matthias; Conrad, Stefan: Exploring the Effects of Cross-Genre Machine Learning for Author Profiling in PAN 2016. In: Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum,. pp. 970–977, 2016.

[MT13] Mohammad, Saif M.; Turney, Peter D.: Crowdsourcing a Word-Emotion Association Lexicon. 29(3):436–465, 2013.

[Na16] Nakov, Preslav; Ritter, Alan; Rosenthal, Sara; Sebastiani, Fabrizio; Stoyanov, Veselin: SemEval-2016 Task 4: Sentiment Analysis in Twitter. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). Association for Computational Linguistics, pp. 1–18, 2016.

[OKA10] O'Connor, Brendan; Krieger, Michel; Ahn, David: TweetMotif: Exploratory Search and Topic Summarization for Twitter. In: Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010. The AAAI Press, pp. 384–385, 2010.

[Pl80] Plutchik, Robert: A general psychoevolutionary theory of emotion. Emotion: Theory, Research, and Experience, 1:3–31, 1980.

[PLV02] Pang, Bo; Lee, Lillian; Vaithyanathan, Shivakumar: Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10. EMNLP '02. Association for Computational Linguistics, pp. 79–86, 2002.

[PSM14] Pennington, Jeffrey; Socher, Richard; Manning, Christopher D.: GloVe: Global Vectors for Word Representation. In: Empirical Methods in Natural Language Processing (EMNLP). pp. 1532–1543, 2014.

[Ra13] Rangel, Francisco; Rosso, Paolo; Koppel, Moshe; Stamatatos, Efstathios; Inches, Giacomo: Overview of the Author Profiling Task at PAN 2013. In: Working Notes for CLEF 2013 Conference. CLEF, 2013.