The Development of a Temporal Information Dictionary for Social Media Analytics

Short Paper

Alivelu Mukkamala
IT University of Copenhagen
Copenhagen S, Denmark
ammu@itu.dk

Roman Beck
IT University of Copenhagen
Copenhagen S, Denmark
romb@itu.dk

Abstract

Dictionaries have been used to analyse text even before the emergence of social media and the use of dictionaries for sentiment analysis there. While dictionaries have been used to understand the tonality of text, so far it has not been possible to automatically detect if the tonality refers to the present, past, or future. In this research, we develop a dictionary containing time-indicating words in a wordlist (T-wordlist). To test how the dictionary performs, we apply our T-wordlist on different disaster related social media datasets. Subsequently we will validate the wordlist and results by a manual content analysis. So far, in this research-in-progress, we were able to develop a first dictionary and will also provide some initial insight into the performance of our wordlist.

Keywords: Method, Social media, Dictionary, Temporal information

Introduction

Since the introduction of social media, unstructured, user-generated content has been created at an unprecedented scale, which became an important source of information for researcher in social science (Thelwall et al. 2008). The predominant research interest has been on analyzing the textual content of social media messages, which contain people's opinions, expectations, feelings, and so on. Different scientific disciplines have used different approaches to analyze text such as manual and automated content analysis, information retrieval, or natural language processing, to mention a few (Loughran and McDonald 2011). The research area that gained most prominence is possibly the sentiment analysis (Read 2005), which allows the automated assessment of positive or negative feelings based on the tonality of a message.

The term sentiment analysis, sometimes referred to as sentiment classification (Read 2005; Read and Carroll 2009), opinion mining (Pang and Lee 2008; Thelwall et al. 2011; Thelwall et al. 2010), subjectivity analysis (Riloff and Wiebe 2003), or polarity classification (Read and Carroll 2009) describes more or less the same concept of carrying out a tonality analysis of a more positive or more negative tone used in a message (Liu 2012). The general assumption is that in general people have certain opinions on topics, and these opinions can be capture by a sentiment analysis (Kim and Hovy 2004). Hence, sentiment analysis is mainly focused at automatically extracting and summarizing the people's opinions about various topics by classifying them (Beigi et al. 2016). Being able to detect the sentiment of a message has been found to be important: For example, by applying sentiment analysis, online retailers can get aggregated results while summarizing the collected reviews with an average sentiment score. In the financial domain, stock market experts can predict stock price fluctuations, based on average sentiments (Read and Carroll 2009). During disasters, a sentiment analysis can help to understand feelings, concerns, or even panic of affected people (Beigi et al. 2016).

While sentiment analysis has illustrated its usefulness, an important dimension so far has been not taken into consideration, and that is to which time dimention the message is referring to. A tweet talking about how much someone loved his first BMW would result in a positive sentiment towards BMW, although the writer of this tweet might not be inclined to buy one nowadays. The same is true for statements such as "IBM was a great company": most sentiment analyses approaches would interpret this as a positive sentiment towards IBM, ignoring the fact that the author might have changed his sentiment in the meanwhile, making such a statement less valuable to build upon a buying decision for IBM stocks.

Time plays also a role in another dimension, namely the need to analyze social media data in close to real time, for example to be able to prioritize emergency response forces along the urgencies of the different needs people in a disaster are suffering from. For example, tweets containing words like "just", "urgent" or "immediately" might help disaster relief forces to prioritize where to move in and help first, if they can analyze the data quick enough, meaning in an automated way, as the following two tweets illustrate.

"A massive earthquake just hit Everest. Basecamp has been severely damaged. Our team is caught in camp 1. Please pray for everyone".

"Urgent. Need 1000 water packets. Please contact us immediately".

Thus, we claim that detecting time-indicating statements in social media is important to refine, for example if a positive sentiment is related to the past, the presence, or the future. It is also important to differentiate what is more urgent in case of an emergency, which also requires an automated detection of time-indicating statements. So far, there has been only limited research that focused on extracting time-indicating statements from social media nor has there been a similar effort as in the case of sentiment analysis to develop a dictionary or wordlist, comprising time-relevant words and phrases. Thus, our research focusses on filling this gap by developing a time-indicating dictionary. For that to do so, in a first step we had to develop a "dictionary development methods", since there are many dictionaries in use in sentiment analysis, but it seems there is no structured approach how to develop one. Thus in this paper our research question is:

How can time-indicating expressions be captured in a dictionary to automatically assess social media data in close to real time?

The remaining research in progress paper is organized as follows. In section two, we provide a literature review on prior research on detecting time in textual statement, as well as the dictionaries have been create so far using supervised approaches under data analytics. A detailed description of the methodology we developed to derive a dictionary will be presented in the third section. In section four, we present our preliminary results regarding the development of our time-indication dictionary and the data sources we used to extract our time-indicating expressions. Based on preliminary results, in the last section we discuss our future research where we explain what we intend to do in order to make our wordlist more robust and rigorous to extract more appropriate data from the given content so that it can be generalizable to other domains as well.

Literature review

Temporal Information

The temporal (time) information about events embedded in different types of text is of utmost importance to understand exact time or time period the text is referring to (Alonso et al. 2007) or to answer the questions in a given news article regarding events (Pustejovsky et al. 2003a). Since time is playing an important dimension, extraction and normalization of temporal information has been considered as an essencial task. For this purpose, temporal information extraction and retrieval gained attention back (Allen 1983) where James F. Allen proposed an interval-based temporal logic to reason about temporal knowledge and temporal intervals using the computational approach. In general, one can understand and group the temporal information into four categories: date, time, duration, and set (Pustejovsky et al. 2005) as discussed further. In contrast to time and date expressions which provide specific information about a point in time, duration expressions mainly give information about the length of an interval (Strötgen and Gertz 2013). Furthermore, according to the study of (Schilder and Habel 2001), time denoting expressions in a document come in three different types: explicit reference, index reference, and vague reference. Date expressions such as “18.08.71” provide an explicit reference and point to a precise moment and thus can be normalized easily. As part of indexical reference, temporal expressions (such as “today”, “by last week” etc.) can only be evaluated via the presence of a time stamp in the document. Other types of temporal expressions (such as “in several weeks”, “in the evening”, etc.) express vague temporal information that is difficult to place on a timeline. However, other studies refer to indexical references as relative expressions, where it is argued that context information such as document creation time or another temporal information is necessary to normalize the temporal expression in the documents (Alonso et al. 2011; Strötgen and Gertz 2013). Moreover, implicit expressions such as names of holidays (Christmas 2016) and events can be normalized by their temporal semantics. “The normalization task of a temporal tagger is to assign the same value to all expressions carrying the same semantics or referring to the same point in time” (Strötgen and Gertz 2013).

In continuation of the above discussion, there has been a great amount of interest from the Natural Language Processing (NLP) community in extracting temporal relationships and events from textual corpura such as news media and other formally written texts. The main focus of research in this direction is to identify temporal events and expressions from documents and to establish temporal relationships between such time events and time-dependent facts. One of the first research initiatives, TIMEX2 (Ferro et al. 2001) a standard for annotation of temporal expressions was initially developed as part of TIDES (Translingual Information Detection, Extraction, and Summarization) program. Based on the TIDES TIMEX2 annotation effort, TimeML (Pustejovsky et al. 2003a), a temporal markup language and Timebank corpus (Pustejovsky et al. 2003b) containing annotated events, times and temporal relations was developed to identify events and temporal expressions in natural language texts. After several iterations, TimeML language has become a gold standard for annotating temporal information (Verhagen et al. 2007). Several research groups have also developed tools and toolkits for performing temporal analysis on texts. Based on Timex3 (Group 2009) annotation standards for temporal information, temporal taggers like HeidelTime (Strötgen and Gertz 2010), Stanford temporal tagger (Chang and Manning 2012) were developed to recognize and normalize the time expressions in the textual documents. Stanford temporal tagger is a rule based tagger that is built on top of Stanford POS-tagger and named-entity taggers and offers a good accuracy in identifying the temporal expressions from text.

Moreover, as part of semantic evaluation initiative from NLP community (SemEval 2007-2017), the tasks of time annotation has received greater interest among the NLP community (Group 2009; UzZaman et al. 2012; Verhagen et al. 2007; Verhagen et al. 2010) for evaluating time expressions, events, and temporal relations among the multiple languages. To detect features of time series of facts from a large number of documents, techniques such as joint inference for temporal scoping was used (Ling and Weld 2010; Talukdar et al. 2012). To enhance summarization of multi documents, the study (Ng et al. 2014) focused on the temporal information in the form of timelines. In order to generate a timeline, an automated processing system was employed, through which three features were derived to measure and recognize the importance of sentences. To overcome any potential errors from underlying temporal processing system, a reliability filtering metric was used to decide when the important temporal information should be used. A multi-document summarization is useful where the event occurrences happen in a chronological order (Mani and Wilson 2000). To extract temporal information, in addition to news articles (Ferro et al. 2005; Pustejovsky et al. 2003a), Wikipedia documents (Strötgen and Gertz 2013), and scientific documents (Strötgen and Gertz 2012) were analyzed. To some extent, informal discussions of online communities were also used, to tag, retrieve and normalize temporal information (Wen et al. 2013). Most importantly the challenges in extracting temporal orientation from social media messages such as Facebook messages (Schwartz et al. 2015) were explored and discussed.

Distinct from the previous research, a temporal ontology, TempoWordNet (Ga et al. 2014; Hasanuzzaman et al. 2016) was constructed automatically by adding temporal information to the words from WordNet (Miller 1995) using a two-step classification approach. Using the similar classification approach, another research work (Kolomiyets et al. 2011) explored the task of recognizing time expressions using a number of bootstrapping strategies to generate additional training set documents that are supplemented with temporal words taken from WordNet (Miller 1995) and Latent Word Language Model (Deschacht and Moens 2009). However, we argue that if the time related words are extracted automatically from the the WordNet, these words are not representative of the words used by ordinary people in their daily communications. Hence, in our initial step, we prefer to construct the dictionary by manually collecting temporal words, and then later on we compare and add words from other sources.

Our approach differs in two aspects when compare with all the above mentioned NLP methods. First, most of the above mentioned methods use advanced NLP techniques such as parsing, classification etc, to identify and extract temporal events and thereby to find temporal relationships between the events. On the contrary, our approach uses a simple, easy lexicon-based approach using manually collected time words to identify and filter the texts containing time-indicating information. Second, most of the above mentioned temporal work primarily targeted for news media and articles (such as Wikipedia) where the language styles are generally formal, hence as also indicated in (Wen et al. 2013), applying these techniques to more informally written texts such as social media posts is challenging. As our primary focus is to identify time-indicating expressions in social media texts, we resorted to a lexicon-based approach that is suitable for processing of social media texts in close to real-time.

Sentiment Analysis and Dictionaries

In general, in order to perform sentiment analysis, research focused on two types of approaches which are unsupervised, dictionary-based approaches (Abdulla et al. 2016; Backfried and Shalunts 2016; Taboada et al. 2011; Thelwall et al. 2011; Thelwall et al. 2010), and supervised, machine learning approaches (Abbasi et al. 2008; Gonçalves et al. 2013; Pang and Lee 2008; Read 2005; Yang et al. 2010). The latter approaches are used to build the classifiers, where manually labelled data is used as training set for a supervised machine learning approach. It is argued that classifiers give high accurate results in detecting the sentiment and polarity of a given text (Boiy et al. 2007; Chaovalit and Zhou 2005). However, classifiers are domain specific, hence when applied in another context its' performance drops considerably (Aue and Gamon 2005; Taboada et al. 2011). A recent study classified the sentiment analysis approaches on a much broader spectrum by adding unsupervised, semi-supervised and hybrid approaches, in addition to the approaches mentioned above (Madhoushi et al. 2015).

In a dictionary-based approach, sentiment scores are assigned to a list of words to measure either semantic orientation or polarity (positive or negative), or strength (valence) of a given text (Nielsen 2011; Taboada et al. 2011)). Constructing a sentiment dictionary manually is labor intense and time consuming hence most of the sentiment analysis research depends on preexisting, manually constructed dictionaries.

For example, the Linguistic Inquiry and Word Count (LIWC) application consists of an internal dictionary which was compiled manually (Pennebaker et al. 2001). It consists of 4,500 words and word stems where words initially were collected from different sources while more words were added to it over the time. The words were grouped into, emotion or affective subcategories and arranged hierarchically. Altogether, there are around 66 categories with varying numbers of words where the words are categorized either into positive or negative. In recent versions, the dictionary was updated with functional words, to name a few, conjunctions, adverbs, quantifiers, or commonly used verbs. Moreover, some original categories were removed as they are not used so often (Pennebaker et al. 2001). Another lexicon, Affective Norms for English Words (ANEW) has 2,477 unique words where the words are scored for valency that range from −5 (very negative) to +5 (very positive) (Bradley and Lang 1999). A few other widely used dictionaries in sentiment analysis are Opinion lexicon (Wilson et al. 2005), SentiWordNet (Esuli and Sebastiani 2007), WordNet (Miller et al. 1990), WordNet-Affect (Esuli and Sebastiani 2007), and AFINN-96 (Nielsen 2011). There are also certain domain specific dictionaries (Olteanu et al. 2014; Temnikova et al. 2015) and language specific dictionaries (Madhoushi et al. 2015) as well. Even though there are many dictionaries, it is still unclear how to build a dictionary in the best possible way (Nielsen 2011) and there are no standardized procedures or commonly accepted methods in place how to build them (Deng et al. 2017). Unlike the sentiment analysis dictionaries as mentioned previously, there are no existing dictionaries available for time-related information.

Methodology-how we developed our dictionary

To achieve our research objectives, we followed a methodology that consists of three phases as shown in Figure 1. The first phase comprises the building of a T-wordlist by collecting the words representing time. The second phase consists of data collection and pre-processing of social media data and applying the T-wordlist on it. Through the use of a manual content analysis, the third phase primarily focusses on the validation of the data extracted by the application of T-wordlist.

Regarding alternative approaches to our methodology, we could have chosen different approaches such as classification based methods using supervised machine learning, temporal taggers using advanced NLP methods such as Stanford Temporal Tagger (Chang and Manning 2012) etc. But since our focus is on developing a simple and transparent methodology that is easy to understand and adopt by disaster and other relief organizations, we rather focused on developing our T-wordlist using the lexicon-based approach using the manually collected words indicating time information. Moreover, our chosen lexicon-based approach is also inline with the methodology of CrisisLex lexicon (Olteanu et al. 2014) and in fact our methodology works very well in conjunction with CrisisLex lexicon by complementing it to identify time-indicating expressions from crisis-related messages during disasters.

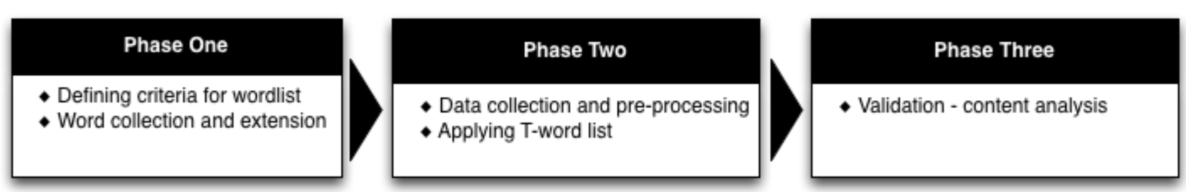

Figure 1. T-wordlist building process

The criteria for a time-indicating wordlist

In order to answer our research question, as a first step we developed a preliminary version of a wordlist containing time indicating and representing words (time indicating words) such as *after, soon, tomorrow, later*, which can extract temporal information from the data. Since there are no existing dictionaries available we had to start from the scratch to develop our T-wordlist. The selection of words defining our T-wordlist involved several iterations. The initial idea was to collect a group of words that explains the temporal information from the disaster social media data. However, over time we expanded the T-wordlist considerably by defining the criteria and adding more words as explained below.

Step 1. Defining criteria: In the development of the T- wordlist, most important requirement is to consider the right words and phrases that contain time information. So firstly, we focused on inclusion and exclusion of words. After several discussions, both the authors decided and defined the criteria for time relevant words as follows. The first criteria: we included words which indicate direct temporal expressions such as *years ago*, *last month*, *tomorrow* and so on. Second criteria: words which signal and help in interpretation of temporal expressions such as temporal prepositions (such as *during on*, *at*, *for*) and connectives (such as *before*, *after*, *while*) (Pustejovsky et al. 2003b). Third criteria: key time words that indirectly signal the time information such as *immediately*, *urgent*, *now*, and so on. Our focus was to analyse data from social media during times of disaster data and not news articles or text summarization (Chambers et al. 2007). Therefore, we decided to exclude the tens verbs such as *has left, was affected*, etc., which indicate the event expressions (Pustejovsky et al. 2003b). Once we defined our criteria, we proceeded to the next step of collecting words from an online dictionary.

Step 2. Word collection and extension: As mentioned previously, unlike dictionaries for sentiment analysis (Esuli and Sebastiani 2007; Nielsen 2011; Strapparava and Valitutti 2004) and domain specific dictionaries (Abdulla et al. 2016; Loughran and McDonald 2011) which are mostly built on already existing dictionaries (Miller 1995; Miller et al. 1990; Stone and Hunt 1963), there are no dictionaries available specifically pertaining to time words only. Therefore, we had to rely on online dictionaries to collect the words and used dictionary.com and Oxford dictionary as our main sources to collect the time relevant words. Initially, we collected only a few common words that provide time information. Later on we extended our wordlist by looking at synonyms and antonyms. In this process, we also used thesauruses to collect more words and phrases. Altogether our initial set of T-wordlist contains around 476 words. The T-wordlist does not only consist unigrams but also phrases.

Once the T-wordlist is now ready, so in step 3 we apply it on disaster-related social media data to extract the messages with time relevant information.

Data collection, preprocessing and application of T-wordlist

In this phase, we applied the T-wordlist on disaster social media data. The social media data that we focused on and collected was the Twitter data from both manmade and natural disasters, such as wild fires, bush fire, floods, shootings, earthquake and so on. Moreover, these disasters occurred in different countries across the world, e.g., Colorado, Philippines, Australia, Singapore, Los Angeles and so on. We will explain in detail the data collection and pre-processing in the following section. We collected publicly available 12 different disaster datasets (Olteanu. 2017) each consisting of approximately 1,000 to 1,200 tweets from different disasters such as wild fires, floods, shootings and so on. The detail information of datasets is provided below Table 1. All together the datasets consist of 12,831 tweets from disparate datasets.

After collecting the data, we randomly took a sample data from each dataset to examine whether or not the dataset consists tweets in English language only. While performing this task, we noticed there are few tweets from languages other than English. For example, in the datasets of Philipinnes floods and typhoon Pablo. Hence we realised pre-processing is an important and required step to proceed to the further analysis. In order to clean and reduce the noise in the data, we pre-processed each dataset to eliminate the tweets other than English language. Of the 12 datasets, we identified all the datasets have language other than English. The results are presented in Table 1.

Table 1. Data Description

Name of dataset/Year	Total number of tweets	English Tweets	Non-English Tweets
Colorado wildfires – 2012	1,200	1,163	37
Philipinnes floods - 2012	1,000	851	149

Typhoon Pablo - 2012	1,000	821	179
Australia bushfire - 2013	1,200	1,174	26
Bohol earthquake - 2013	1,000	837	163
Brazil nightclub fire - 2013	1,000	407	593
Glasgow helicopter crash - 2013	1,100	1,079	21
Los Angeles airport shootings - 2013	1,032	938	94
New York train crash - 2013	1,000	968	32
Savar building collapse -2013	1,250	1,107	143
Singapore haze - 2013	1,000	947	53
Typhoon Yolanda - 2013	1,049	946	103

Table 1. Data Description

Later, we applied the T-wordlist on the individual datasets. We segregated the tweets into two categories. The tweets that are matching with words of our T-wordlist and the tweets that are no way related or containing any of the words from our T-wordlist. Around 45% of T- words represented in the data, will be discussed in the results section. All together we extracted 4,791 number of tweets (43% of total tweets) as presented in Table 2Table 1. At this stage, to ensure whether or not the data is representing temporal information, validation is important.

Validation of the T-wordlist

Data validation is important part of any study because previous studies mentioned that the same word provides different meaning in different context. In order to ascertain whether the tweet is accurate enough in representing the time relevant information as we intended, we conducted a manual content analysis. In general, on an aggregated level to identify meaningful insights out of data, and to obtain replicable, reliable and valid inferences, researchers often apply a content analysis (Krippendorff 1989). This approach is often used in IS research either to figure out the categories inductively or to classify the information based on pre-defined categories. However, in our current study we neither wanted to categorise the data nor did we classify new information: as mentioned earlier, we wanted to check whether or not a tweet is containing time-relevant information that one can extract using the T-wordlist. Both researchers intensively worked on time-relevant words and discussed rigorously what constitutes temporal information and what does not. However, we felt a pilot study is required before we analyse a sample from all the twelve datasets. For this, we randomly selected a small sample of tweets from four datasets. In the pilot study, one of the researchers analysed a sample dataset individually to assess whether or not tweets are representing temporal information. Later, both researchers looked into the results and discussed them. Again, to ensure validity, in the subsequent stage, both researchers analysed a sample data of around 1,000 tweets from the matched tweets dataset representing different types of disasters. The preliminary results are discussed in the following section.

In the future, as part of a further evaluation of our T-wordlist and also to bench mark it against established toolkits such as TempoWordNet (Ga et al. 2014), we will apply both the T-wordlist and other toolkits on the disaster social media datasets from CrisisLex.org (Olteanu. 2017) and also on the data sets collected by us (such as Chennai floods Twitter data). In order to evaluate the results, we will also employ services from crowdsourcing platforms such as MicroWorker, Mechanical Turk etc. and in this way, we will compare the results and bench mark our T-wordlist against the existing methods and toolkits. After the benchmarking process, we will also enrich the T-wordlist by supplementing with suitable temporal expression words taken from TempoWordNet (Ga et al. 2014) and other relevant sources to make it more robust and useful.

Results and future work

As base of our research in progress study, we used our T-wordlist which consists of 470 time-indicating phrases and a Twitter dataset consisting of 11,238 tweets in English language, as has been shown in Table 1. After applying the T-wordlist we extracted the tweets consisting time indicating phrases (Table 2). Only 210 words from the T-wordlist matched with words and phrases in our Twitter dataset. The frequency of words and phrases ranged from a single match to 628 matches in the 4791 tweets that were matched by the T-wordlist. Furthermore, 87 words of our T-wordlist appeared more than 10 times in the matched Twitter dataset. As a next step, we have to cross-check the data to make sure whether the tweet really contains the time-related information and also have to have a look into the tweets which had no time-related information, to check if we have missed important phrases or words.

Table 2. Matching of T-words

Name of dataset/Year	English Tweets	Match (T-word)	Non-Match
Colorado wildfires - 2012	1,163	490	673
Philipinnes floods - 2012	851	307	544
Typhoon Pablo - 2012	821	340	481
Australia bushfire - 2013	1,174	522	652
Bohol earthquake - 2013	837	331	506
Brazil nightclub fire - 2013	407	198	209
Glasgow helicopter crash - 2013	1,079	441	638
Los Angeles airport shootings - 2013	938	450	488
New York train crash - 2013	968	414	554
Savar building collapse -2013	1,107	407	700
Singapore haze - 2013	947	505	442
Typhoon Yolanda - 2013	946	386	560

Table 2. Matching of T-words

Based on our preliminary results, in our future research we mainly will focus on compiling a more complete T-wordlist to achieve more accuracy. To make our wordlist more rigorous and robust, we have to tackle the problem how to deal with changing meanings of words, depending on the context in which they are used. For example, prepositions like, *on* (Monday), *in* (the morning), *at* (night), *by* (11 o'çlock) indicate temporal information. However, the same prepositions might also indicate the position and direction in different context. For example, *in* (the picture), *on* (the left), *at* (a concert), (standing) *by*. In this regard, by taking advantage of the support of one of the crowd sourcing platforms, we plan to make our wordlist more useful and generalizable to use it for different data analyses. Furthermore, we will use the services from crowd sourcing platforms to categorise the words (based on meaning) into past/present/future categories and also to benchmark out T-wordlist against existing temporal tools and toolkit as explained in the validation section.

References

Abbasi, A., Chen, H., and Salem, A. 2008. "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," *ACM Transactions on Information Systems (TOIS)* (26:3), p. 12.

Abdulla, N. A., Ahmed, N. A., Shehab, M. A., Al-Ayyoub, M., Al-Kabi, M. N., and Al-rifai, S. 2016. "Towards Improving the Lexicon-Based Approach for Arabic Sentiment Analysis," in *Big Data: Concepts, Methodologies, Tools, and Applications*. IGI Global, pp. 1970-1986.

Allen, J. F. 1983. "Maintaining Knowledge About Temporal Intervals," *Communications of the ACM* (26:11), pp. 832-843.

Alonso, O., Gertz, M., and Baeza-Yates, R. 2007. "On the Value of Temporal Information in Information Retrieval," *ACM SIGIR Forum*: ACM, pp. 35-41.

Alonso, O., Strötgen, J., Baeza-Yates, R. A., and Gertz, M. 2011. "Temporal Information Retrieval: Challenges and Opportunities," *Twaw* (11), pp. 1-8.

Aue, A., and Gamon, M. 2005. "Customizing Sentiment Classifiers to New Domains: A Case Study," *Proceedings of recent advances in natural language processing (RANLP)*: Citeseer, pp. 2-1.

Backfried, G., and Shalunts, G. 2016. "Sentiment Analysis of Media in German on the Refugee Crisis in Europe," *International Conference on Information Systems for Crisis Response and Management in Mediterranean Countries*: Springer, pp. 234-241.

Beigi, G., Hu, X., Maciejewski, R., and Liu, H. 2016. "An Overview of Sentiment Analysis in Social Media and Its Applications in Disaster Relief," in *Sentiment Analysis and Ontology Engineering*. Springer, pp. 313-340.

Boiy, E., Hens, P., Deschacht, K., and Moens, M.-F. 2007. "Automatic Sentiment Analysis in on-Line Text," *ELPUB*, pp. 349-360.

Bradley, M. M., and Lang, P. J. 1999. "Affective Norms for English Words (Anew): Instruction Manual and Affective Ratings," Technical report C-1, the center for research in psychophysiology, University of Florida.

Chambers, N., Wang, S., and Jurafsky, D. 2007. "Classifying Temporal Relations between Events," *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*: Association for Computational Linguistics, pp. 173-176.

Chang, A. X., and Manning, C. D. 2012. "Sutime: A Library for Recognizing and Normalizing Time Expressions," *LREC*, pp. 3735-3740.

Chaovalit, P., and Zhou, L. 2005. "Movie Review Mining: A Comparison between Supervised and Unsupervised Classification Approaches," *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*: IEEE, pp. 112c-112c.

Deng, Q., Hine, M., Ji, S., and Sur, S. 2017. "Building an Environmental Sustainability Dictionary for the It Industry," *Proceedings of the 50th Hawaii International Conference on System Sciences*.

Deschacht, K., and Moens, M.-F. 2009. "Semi-Supervised Semantic Role Labeling Using the Latent Words Language Model," in: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*. Singapore: Association for Computational Linguistics, pp. 21-29.

Esuli, A., and Sebastiani, F. 2007. "Sentiwordnet: A High-Coverage Lexical Resource for Opinion Mining," *Evaluation*), pp. 1-26.

Ferro, L., Gerber, L., Mani, I., Sundheim, B., and Wilson, G. 2005. "Tides 2005 Standard for the Annotation of Temporal Expressions,").

Ferro, L., Mani, I., Sundheim, B., and Wilson, G. 2001. "Tides Temporal Annotation Guidelines-Version 1.0. 2," *The MITRE Corporation, McLean-VG-USA*).

Ga, #235, Dias, l. H., Hasanuzzaman, M., St, #233, Ferrari, p., and Mathet, Y. 2014. "Tempowordnet for Sentence Time Tagging," in: *Proceedings of the 23rd International Conference on World Wide Web*. Seoul, Korea: ACM, pp. 833-838.

Gonçalves, P., Araújo, M., Benevenuto, F., and Cha, M. 2013. "Comparing and Combining Sentiment Analysis Methods," *Proceedings of the first ACM conference on Online social networks*: ACM, pp. 27-38.

Group, T. W. 2009. "Guidelines for Temporal Expression Annotation for English for Tempeval 2010."

Hasanuzzaman, M., Dias, G., Ferrari, S., and Mathet, Y. 2016. "Iterative Tempowordnet," in *The Semantic Web: Eswc 2016 Satellite Events, Heraklion, Crete, Greece, May 29 – June 2, 2016, Revised Selected Papers,* H. Sack, G. Rizzo, N. Steinmetz, D. Mladenić, S. Auer and C. Lange (eds.). Cham: Springer International Publishing, pp. 40-45.

Kim, S.-M., and Hovy, E. 2004. "Determining the Sentiment of Opinions," *Proceedings of the 20th international conference on Computational Linguistics*: Association for Computational Linguistics, p. 1367.

Kolomiyets, O., Bethard, S., and Moens, M.-F. 2011. "Model-Portability Experiments for Textual Temporal Analysis," *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*: Association for Computational Linguistics, pp. 271-276.

Krippendorff, K. 1989. "Content Analysis," in *International Encyclopedia of Communication,* G.G. E. Barnouw, W. Schramm, T. L. Worth, & L. Gross (ed.). pp. 403-407.

Ling, X., and Weld, D. S. 2010. "Temporal Information Extraction," *AAAI*, pp. 1385-1390.

Liu, B. 2012. "Sentiment Analysis and Opinion Mining," *Synthesis lectures on human language technologies* (5:1), pp. 1-167.

Loughran, T., and McDonald, B. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance* (66:1), pp. 35-65.

Madhoushi, Z., Hamdan, A. R., and Zainudin, S. 2015. "Sentiment Analysis Techniques in Recent Works," *Science and Information Conference (SAI), 2015*: IEEE, pp. 288-291.

Mani, I., and Wilson, G. 2000. "Robust Temporal Processing of News," *Proceedings of the 38th annual meeting on Association for Computational Linguistics*: Association for Computational Linguistics, pp. 69-76.

Miller, G. A. 1995. "Wordnet: A Lexical Database for English," *Communications of the ACM* (38:11), pp. 39-41.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. J. 1990. "Introduction to Wordnet: An on-Line Lexical Database," *International journal of lexicography* (3:4), pp. 235-244.

Ng, J.-P., Chen, Y., Kan, M.-Y., and Li, Z. 2014. "Exploiting Timelines to Enhance Multi-Document Summarization," *ACL (1)*, pp. 923-933.

Nielsen, F. Å. 2011. "A New Anew: Evaluation of a Word List for Sentiment Analysis in Microblogs," *arXiv preprint arXiv:1103.2903*).

Olteanu, A., Castillo, C., Diaz, F., and Vieweg, S. 2014. "Crisislex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises," *ICWSM*.

Olteanu., A. 2017. "Crisislex.Org." from http://crisislex.org/

Pang, B., and Lee, L. 2008. "Opinion Mining and Sentiment Analysis," *Foundations and trends in information retrieval* (2:1-2), pp. 1-135.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. 2001. "Linguistic Inquiry and Word Count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates* (71:2001), p. 2001.

Pustejovsky, J., Castano, J. M., Ingria, R., Sauri, R., Gaizauskas, R. J., Setzer, A., Katz, G., and Radev, D. R. 2003a. "Timeml: Robust Specification of Event and Temporal Expressions in Text," *New directions in question answering* (3), pp. 28-34.

Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., and Ferro, L. 2003b. "The Timebank Corpus," *Corpus linguistics*, p. 40.

Pustejovsky, J., Knippen, R., Littman, J., and Saurí, R. 2005. "Temporal and Event Information in Natural Language Text," *Language resources and evaluation* (39:2), pp. 123-164.

Read, J. 2005. "Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentiment Classification," *Proceedings of the ACL student research workshop*: Association for Computational Linguistics, pp. 43-48.

Read, J., and Carroll, J. 2009. "Weakly Supervised Techniques for Domain-Independent Sentiment Classification," *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*: ACM, pp. 45-52.

Riloff, E., and Wiebe, J. 2003. "Learning Extraction Patterns for Subjective Expressions," *Proceedings of the 2003 conference on Empirical methods in natural language processing*: Association for Computational Linguistics, pp. 105-112.

Schilder, F., and Habel, C. 2001. "From Temporal Expressions to Temporal Information: Semantic Tagging of News Messages," *Proceedings of the workshop on Temporal and spatial information processing-Volume 13*: Association for Computational Linguistics, p. 9.

Schwartz, H. A., Park, G., Sap, M., Weingarten, E., Eichstaedt, J., Kern, M., Stillwell, D., Kosinski, M., Berger, J., and Seligman, M. 2015. "Extracting Human Temporal Orientation from Facebook Language," *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 409-419.

SemEval. 2007-2017. "International Workshop on Semantic Evaluation." 2017

Stone, P. J., and Hunt, E. B. 1963. "A Computer Approach to Content Analysis: Studies Using the General Inquirer System," *Proceedings of the May 21-23, 1963, spring joint computer conference*: ACM, pp. 241-256.

Strapparava, C., and Valitutti, A. 2004. "Wordnet Affect: An Affective Extension of Wordnet," *LREC*: Citeseer, pp. 1083-1086.

Strötgen, J., and Gertz, M. 2010. "Heideltime: High Quality Rule-Based Extraction and Normalization of Temporal Expressions," *Proceedings of the 5th International Workshop on Semantic Evaluation*: Association for Computational Linguistics, pp. 321-324.

Strötgen, J., and Gertz, M. 2012. "Temporal Tagging on Different Domains: Challenges, Strategies, and Gold Standards," *LREC*, pp. 3746-3753.

Strötgen, J., and Gertz, M. 2013. "Multilingual and Cross-Domain Temporal Tagging," *Language Resources and Evaluation* (47:2), pp. 269-298.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. 2011. "Lexicon-Based Methods for Sentiment Analysis," *Computational linguistics* (37:2), pp. 267-307.

Talukdar, P. P., Wijaya, D., and Mitchell, T. 2012. "Coupled Temporal Scoping of Relational Facts," *Proceedings of the fifth ACM international conference on Web search and data mining*: ACM, pp. 73-82.

Temnikova, I., Castillo, C., and Vieweg, S. 2015. "Emterms 1.0: A Terminological Resource for Crisis Tweets," *ISCRAM 2015 proceedings of the 12th international conference on information systems for crisis response and management*.

Thelwall, M., Buckley, K., and Paltoglou, G. 2011. "Sentiment in Twitter Events," *Journal of the American Society for Information Science and Technology* (62:2), pp. 406-418.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. 2010. "Sentiment Strength Detection in Short Informal Text," *Journal of the American Society for Information Science and Technology* (61:12), pp. 2544-2558.

Thelwall, M., Wouters, P., and Fry, J. 2008. "Information-Centered Research for Large-Scale Analyses of New Information Sources," *Journal of the American Society for Information Science and Technology* (59:9), pp. 1523-1527.

UzZaman, N., Llorens, H., Allen, J., Derczynski, L., Verhagen, M., and Pustejovsky, J. 2012. "Tempeval-3: Evaluating Events, Time Expressions, and Temporal Relations," *arXiv preprint arXiv:1206.5333*).

Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G., and Pustejovsky, J. 2007. "Semeval-2007 Task 15: Tempeval Temporal Relation Identification," *Proceedings of the 4th International Workshop on Semantic Evaluations*: Association for Computational Linguistics, pp. 75-80.

Verhagen, M., Sauri, R., Caselli, T., and Pustejovsky, J. 2010. "Semeval-2010 Task 13: Tempeval-2," *Proceedings of the 5th international workshop on semantic evaluation*: Association for Computational Linguistics, pp. 57-62.

Wen, M., Zheng, Z., Jang, H., Xiang, G., and Rosé, C. P. 2013. "Extracting Events with Informal Temporal References in Personal Histories in Online Communities," *ACL (2)*, pp. 836-842.

Wilson, T., Wiebe, J., and Hoffmann, P. 2005. "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," *Proceedings of the conference on human language technology and empirical methods in natural language processing*: Association for Computational Linguistics, pp. 347-354.

Yang, C. C., Tang, X., Wong, Y., and Wei, C.-P. 2010. "Understanding Online Consumer Review Opinions with Sentiment Analysis Using Machine Learning," *Pacific Asia Journal of the Association for Information Systems* (2:3).