FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

A Social Affective Text Mining Approach for Detecting Human Emotions on Specific Topics in Twitter Data

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Wirtschaftsinformatik

eingereicht von

Alexander Ortner
Matrikelnummer 0925455

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung: Ao.Univ.Prof Dr. Wolfgang Hofkirchner

Wien, 21.08.2014

(Unterschrift Verfasser) (Unterschrift Betreuung)

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

A Social Affective Text Mining Approach for Detecting Human Emotions on Specific Topics in Twitter Data

MASTER'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Business Informatics

by

Alexander Ortner
Registration Number 0925455

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: Ao.Univ.Prof Dr. Wolfgang Hofkirchner

Vienna, 21.08.2014

(Signature of Author) (Signature of Advisor)

Erklärung zur Verfassung der Arbeit

Alexander Ortner
Obkirchergasse 16, 1190 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

(Ort, Datum)	(Unterschrift Verfasser)

Danksagung

Ein großes Dankeschön geht an meine Eltern, Aysin und Günter Ortner, die mir das Studium an der TU und meinen Auslandsaufenthalt in London ermöglicht haben. Ebenfalls sehr wertvolle Unterstützung habe ich von meinen Großeltern, Hermine und Walter Ortner erhalten. Weiters danke ich meinen Betreuern Dr. Joachim Klerx und Ao. Univ. Prof. Dr. Wolfang Hofkirchner für die gute Zusammenarbeit. Joachim, für die interessanten Einblicke in die Welt des Horizon Scanning und das Herausarbeiten der Grundidee. Wolfgang, für die kooperative und unkomplizierte Gestaltung der Zusammenarbeit und die kritischen Fragestellungen, die maßgeblich zur Verbesserung der Arbeit beigetragen haben. Ein besonderer Dank geht an Daniel Kaiser, für die große Hilfe beim Korrekturlesen der Arbeit. Weiters danke ich Matthias Trümmel für das gemeinsame Brainstorming, sowie Thomas Ziegelbecker, Christoph Moser und Damir Dizdarevic für die vielen Stunden, die wir in der Bibliothek verbracht haben.

Acknowledgements

I would like to thank my parents, Aysin und Günter Ortner, who supported me during my studies at TU and my year abroad in London. I also received valuable support from my grandparents, Hermine und Walter Ortner. Furthermore, I would like to express my sincere gratitude to my advisors Dr. Joachim Klerx and Ao. Univ. Prof. Dr. Wolfang Hofkirchner. Joachim, for his elaborations on horizon scanning and the construction of our initial idea. Wolfgang, for his cooperative straightforward collaboration and asking of critical questions, which were extremely beneficial to the continuous improvement of this work. I would also like to thank Daniel Kaiser for his great help in proofreading. Furthermore, I thank Matthias Trümmel for brainstorming with me, as well as Thomas Ziegelbecker, Christoph Moser and Damir Dizdarevic for spending endless hours with me in the library.

Abstract

The increase in availability of public unstructured data has led to an increasing interest in analysing and understanding its contents. In this context, text mining techniques have been developed, most of which classify text with respect to its positive or negative polarity [25] [32]. More sophisticated emotion mining and social affective text mining techniques deal with the detection of specific emotions in text. Horizon scanning is a research field which may benefit a lot from text mining but little work has been done to support this idea. Its aim is to identify weak signals for emerging issues, which is traditionally done through producing a list of topics by manual scanning of text documents [24] [3]. According to current research in horizon scanning, it can be argued that an indicator for topic relevance is the occurrence of emotions in the written context of a specific topic. Based on this assumption, the aim of this work was to design an emotion mining approach for Twitter micro-blogging posts which supports the horizon scanning process. It considers Twitter-specific factors, such as the use of a limited character length and the use of social media language. The proposed approach was evaluated by measuring emotion mining accuracy as well as its applicability to horizon scanning. This was done by using three Twitter corpora: (1) an accuracy evaluation corpus, (2) a corpus containing Tweets from November 2013 to March 2014 with hashtags related to the Ukraine and (3) a corpus containing Tweets from 1 November 2013 posted from UK locations. Precision values of the proposed approach reached an average of *50%* and *two out of four* identified horizon scanning criteria were compatible with the proposed emotion mining approach. These results show that the proposed novel approach is an appropriate tool for emotion mining and horizon scanning based on Twitter data. Future work may aim to increase emotion mining accuracy by performing text dependency parsing and considering factors such as negation and adjectives which are modified by adverbs. Furthermore, Twitter corpus limitations concerning specific locations and hashtags should be altered in order to examine more broadly under which circumstances a corpus may generate usable results for emerging issue identification.

Kurzfassung

Die erhöhte Verfügbarkeit von öffentlich zugänglichen, unstrukturierten Daten hat zu einem gesteigerten Interesse an Verwertungsmöglichkeiten der darin enthaltenen Informationen geführt. In diesem Zusammenhang wurden Text-Mining-Systeme entwickelt, die vorrangig zur Klassifizierung von positiver und negativer Polarität verwendet wurden [25] [32]. Weitreichendere Emotion-Mining-Systeme befassen sich mit der Erkennung von Emotionen in Text. Horizon-Scanning ist ein Forschungsgebiet, welches sehr von automatisierter Textanalyse profitieren könnte, jedoch gibt es wenige Forschungsarbeiten, um dies zu belegen. Das Ziel von Horizon-Scanning ist das Erkennen von schwachen Signalen für Trends, Chancen oder Bedrohungen, was traditionell durch das Erstellen von Themenlisten nach manuellem Untersuchen von Textdokumenten durchgeführt wird [24] [3]. Aktuelle Forschungsarbeiten argumentieren, dass das Auftreten von Emotionen in geschriebener Sprache, im Zusammenhang mit bestimmten Themen, ein Indikator für Themenrelevanz im Horizon-Scanning sein kann. Folgend war das Ziel dieser Arbeit das Erstellen eines Emotion-Mining-Verfahrens für Twitter-Meldungen, welches den Horizon-Scanning-Prozess unterstützt. Es behandelt spezielle Faktoren, wie die eingeschränkte Länge von Twitter-Nachrichten und die Verwendung von Sprache, die in sozialen Medien üblich ist. Das vorgestellte Verfahren wurde auf Genauigkeit in der Identifikation von Emotionen, sowie Anwendbarkeit auf Horizon-Scanning evaluiert. Dafür wurden drei Twitter Korpora verwendet: (1) ein Genauigkeits-Evaluierungs-Korpus, (2) ein Korpus mit Tweets von November 2013 bis März 2013, die im Zusammenhang mit Ukraine-Hashtags stehen und (3) ein Korpus mit Tweets vom 1. November 2013, welche von Orten in Großbritannien versendet wurden. Präzisionswerte des Verfahrens ergaben einen Durchschnitt von *50%* und *zwei von vier* der identifizierten Horizon-Scanning-Kriterien waren kompatibel mit dem vorgestellten Verfahren. Diese Ergebnisse zeigen, dass das vorgeschlagene, neuartige Verfahren geeignet für Horizon-Scanning und Emotion-Mining anhand von Twitter-Daten ist. Zukünftige Forschungsarbeiten könnten die Genauigkeit des Emotion-Mining-Verfahrens verbessern, indem Abhängigkeiten zwischen Satzgliedern und spezielle Eigenschaften wie Verneinungen beachtet werden. Außerdem könnten die Einschränkungen der verwendeten Twitter Korpora bezüglich spezifischer Orte und vorkommender Hashtags verändert werden, um einen breiteren Blick auf die Verwendungsmöglichkeiten von Twitter-Daten für das Erkennen von auftretenden Ereignissen zu bekommen.

Contents

CHAPTER 1

Introduction

1.1 Motivation

Public Unstructured Data The internet has become a place where user behaviour has changed from information consumption to information creation. Forums, blogs and social networks are examples of platforms where people take part in online content-creation. They share facts, opinions and feelings. Most of this is done via text, while video and image sharing have become more popular as increasing internet bandwidth allows quick uploads of media files. As all of this information can be described as unstructured data, information extraction is non-trivial, thus demanding sophisticated methods for gaining insights in underlying semantics. The increase in availability of public unstructured data has led to an increasing interest in analysing and understanding its contents while context-relating the results to various fields such as brand reputation management and product review analysis. In these fields, text mining techniques have been developed, most of which classify text with respect to its positive or negative polarity [25] [32].

Computational Methods for Horizon Scanning The field of horizon scanning may also benefit a lot from text mining but little work has been done to support this idea. The aim of horizon scanning is to identify emerging issues in a given field sufficiently early to conduct research for informing policy and practice [44]. Emerging issues are stories about what the future, or possible futures, may or will look like and are used as common concepts in horizon scanning. Horizon scanning methods for identifying such issues range from simple manual desk research to more sophisticated approaches such as expert surveys and the Delphi method [3]. Recently, research interest in computational text mining methods which facilitate the identification of emerging issues has increased. A potential source of data for this purpose is Twitter, a micro-blogging platform with over 240 million monthly active users that allows posting of 140 character-long short messages [48]. Given the high amount of data produced by Twitter users, extracting indicators for emerging issues from Twitter posts by means of text mining can be approached in various ways. One possible approach is to search for emotional mentions of specific topics. Therefore, emotion mining, a computational technique for identifying emotional bias in text, may be used for emerging issue identification. Emotion Mining deals with the detection of human emotions in written language and is also referred to as social affective text mining.

1.2 Problem Statement

Recently, a lot of research has which describes different ways of computationally classifying text by emotions has been published. A considerable amount of this research has been done to support businesses in reputation management, brand analysis and automatic assessment of product reviews [4]. However, little work has been done that focuses on Twitter data and its unique characteristics, such as a character limit of 140 symbols, the use of social media language and the utilisation of hash tags for topic assignment. Furthermore, existing research neglects aspects that are related to horizon scanning, such as mapping emotions to topics (e.g. {Obama, Love}) and location (e.g. {Texas, Obama, Anger}). Moreover, there is a lack of prototypical implementations and proofs-of-concept of existing frameworks. The main research questions of this work are: (1) *How can you computationally identify emotions concerning specific topics in Twitter micro-blogging posts?* and (2) *Can emotion mining concerning specific topics in Twitter data be used for horizon scanning?*

1.3 Methodological Approach

As the main aim of this work is to design an artifact, the guidelines of the **Design Science** [21] approach were used to derive the following steps:

1. **Analyse existing emotion dictionaries and emotion mining techniques.** Based on current research, benefits and shortcomings of existing techniques for emotion mining and their applicability to the problem space will be analysed.

2. **Design an emotion mining approach for short texts on Twitter tailored to the needs of horizon scanning.** A technical design for emotion mining on Twitter data will be created. Designing an emotion mining approach is the main goal of this work. Trivial methods will be used for dealing with topic and location detection, which are relevant to horizon scanning.

3. **Create a prototype that implements the developed design.** The technical design will be implemented and applied to a Twitter corpus. The prototype will output graphical representations of emotions for specific Twitter topics. A twitter topic is defined by a specific hashtag, for example *#haiti* groups together Tweets related to *Haiti*.

4. **Evaluate emotion mining accuracy and horizon scanning applicability.** For evaluation emotion mining precision, a subset of the Twitter data will be manually classified by human annotators. The result of this classification will be compared to the results computed by the prototype. Furthermore, horizon scanning applicability will be tested by comparing a criteria catalogue to experimental results generated by the implemented prototype.

1.4 Aim and Expected Result

The aim of this work is to design an emotion mining approach which considers the characteristics of Twitter data and outputs emotion mining results in a visual format thus supporting the horizon scanning process. Obvious characteristics of Twitter data are the character limit of 140 symbols, the occurrence of Twitter-specific tags, such as hashtags, and the use of social media language, such as emoticons. A more intricate analysis is done in section 3.1. A visual format that aims to support the horizon scanning process may include some or all of the dimensions: emotion, topics, location and time. One way for visualising these are emotion maps which are geographical maps with additional layers displaying emotions and topics. A more detailed analysis of how to present emotion mining results for horizon scanning purposes is done in section 2.

The expected result consists of the following parts:

(1) A comparison of existing methods and frameworks for emotion mining.

(2) A design for an emotion mining approach tailored to characteristics of Twitter data.

(3) A prototypical implementation that outputs an emotion map.

(4) An evaluation of the implementation, comparing computed results with results of human annotators.

1.5 Structure

Chapter 2 described the state of the art in horizon scanning and detecting emotions in written text with dictionary based techniques. Based on chapter 2 an emotion mining approach is designed in chapter 3 and its implementation is described in chapter 4. In chapter 5, experimental data is used to apply and evaluate the approach. Finally, summary and outlook can be found in chapter 6.

CHAPTER 2

State of the Art in Horizon Scanning and Emotion Mining

The aim of this section is to review current literature in horizon scanning and emotion mining. Section 2.1 gives a brief overview of horizon scanning, focusing on features which are relevant to automated textmining. This is followed by a description of state of the art emotion mining approaches in section 2.2, while section 2.3 describes lexical emotion mining resources that have been proposed by the research community.

2.1 Horizon Scanning

As the aim of this work is to design an emotion mining approach which supports the needs for horizon scanning, this section analyses current research in horizon scanning.

The field of horizon scanning dates back to 1967 when Aguilar [1] described it as a policy tool which aims to investigate novel, unexpected issues and trends. Since then, various other definitions have been proposed. The UK Department for Environment, Food and Rural Affairs [14] defined horizon scanning as "the systematic examination of potential threats, opportunities and likely future developments which are at the margins of current thinking and planning." King and Thomas [24] state that horizon scanning is a "series of workshops, seminars, brainstorming sessions, and other conversations with industry and relevant professional organizations." The aim of horizon scanning is to identify emerging issues in a given field sufficiently early to conduct research for informing policy and practice [44]. These issues can be of political, economical, social, ecological and technical nature [20]. Emerging issues may be related to weak signals which are indicators for long term changes that have not yet become general knowledge [20]. Habegger et al [20] identify five main characteristics of horizon scanning:

- **Viewing and Searching Activities**. Horizon scanning can be done passively by *viewing* information and noticing indicators for emerging issues. In contrast, *searching* is a process where concrete questions are used as starting points for horizon scanners.

- **Mid- or Long-Term Perspective**. Instead of focusing on topics already on the agenda of policy makers, horizon scanners aim to identify future issues.
- **Focus on Margins of Current Thinking**. In order to identify these margins, people, personal networks and various media are potential sources of information.
- **Analysis of External Environments**. New issues in a specific environment may originate from another environment which may seem completely unrelated in the beginning.
- **Knowledge Exchange**. Teams of horizon scanners gather information and make it available to others, e.g. in web applications.

According to Klerx et al. [23], horizon scanning activities lead to results which form a foundation for detecting weak signals. These weak signals can be interpreted as threat or opportunity which is a subjective assessment depending on what perspective the assessor takes.

Amanatidou et al. [3] describe the emergence of issues over time based on the EU FP7 SESTI Project[1] and define criteria for detecting weak signals as listed in table 2.1.

1	Described impact	Is there an impact on society, economy, ecology?
2	Desirability	Which values are addressed in the impact description? What is the narrator's point of view?
3	Factual basis	Are there references to reliable and/or biased sources?
4	Plausible storyline	Is the described scenario plausible?
5	Novelty	Is the issue described new for the policy-makers?
6	Interests at stake	Does the storyline involve strong conflicts or commonalities in interests?
7	Emotional aspects and critical aspects	Does the issue appeal or concern emotional and or ethical legal aspects?
8	Changeability	Can the story or its impact be altered by human action. How does the issue relate to present day decision making and action and on different levels?

Table 2.1: Criteria for Detecting Primary signals [3]

Various methods for detecting weak signals have been employed during EU FP7 SESTI:

- **Manual scanning**. Researcher manually scan research papers and other media.
- **Wiki**. New Wiki entries are used to determine emerging topics and issues.

[1] SESTI - Scanning for Emerging Science and Technology Issues: http://ec.europa.eu/research/fp7/

- **Expert survey**. Companies, think tanks and universities are asked for their opinion.
- **Conferences**. Conferences are assessed for new research topics.
- **Twitter**. Being a micro-blogging platform which is used by various kinds of authors such as politicians, companies and researchers, Twitter is a potential source of data.
- **Text-mining**. As the amount of publicly available information is increasing, automated methods for analysing text have become more important.

Research suggests the use of several other methods, such as morphological analysis, intra- and extrapolation of trends and time series analysis [22]. Morphological analysis is used for structuring a knowledge domain. Trend and time series analysis uses historical data to predict future expectations.

Klerx et al. [23] emphasise that automated horizon scanning relies on sophisticated text mining methods. Examples are meta-data-extraction, topic mining and emotion mining. They further suggest that a first step of horizon scanning should be an automated classification of weak signal detection results. Topics, emotions and other meta-data such as names, organisations, times and locations are relevant to such text analysis programs. Technical details, such as URLs, file names, encoding or domain names and links could also be used. In horizon scanning, time and location are commonly used for structuring existing data. A time dimension enables examination of when text was written and how text content changed over time. For assigning topics and emotions to geographical regions, location information is used. Following the extraction of relevant meta-data, opportunities and threats can be detected by emotional classification of topics.

Horizon Scanning may be seen as part of a larger foresight process. In the last few decades, different foresight studies have been conducted which aim to describe possible futures. However, due to the fact that outcomes of these studies can be deemed irrelevant by conflicting current events, continuous monitoring and analysis of activities is necessary [23]. As depicted in figure 2.1, the foresight process can be divided into three phases: (1) Early Detection, (2) Generating Foresight and (3) Developing Policy Options [20]. Horizon Scanning methods are used in the Early Detection phase.

Some countries, such as the UK, have employed so called horizon scanning centres to facilitate horizon scanning activities. Currently, horizon scanning centers have not employed large scale automated methods [23].

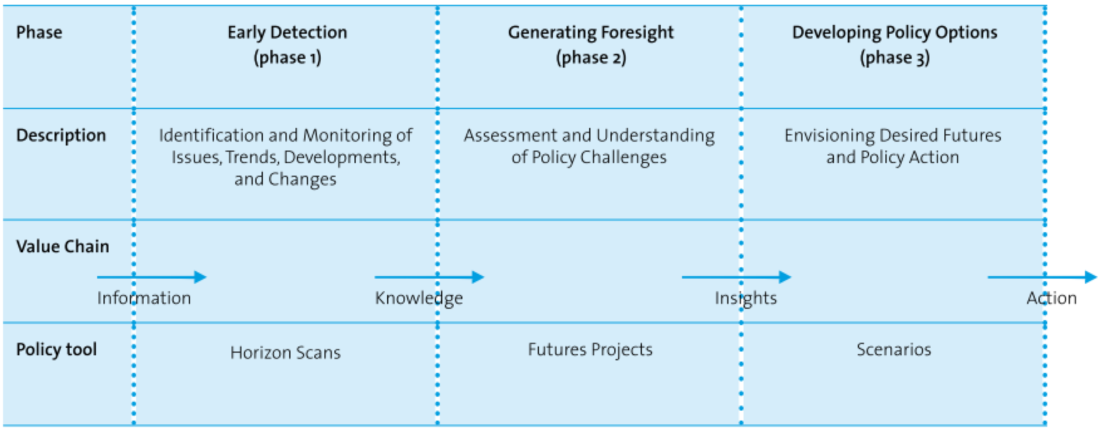


Figure 2.1: Three phases of A foresight process. Illustration taken from [20]

2.1.1 Horizon Scanning Features Relevant to This Work

Based on the analysis of the previous section, a catalogue for indicators of emerging issues can be identified. Indicators are topics with one or more specific characteristics:

1. Novel with increasing popularity
   Topics which newly emerge, leading to an increasing number of people talking about them.

2. Relevant to society
   Topics relevant to society, which are not necessarily on policy agendas.

3. At first single then cross-domain
   Topics which occur in a specific domain and spread to other domains.

4. Weak signals for threats, opportunities, trends
   Topics related to weak signals for threats, opportunities and trends.

This list of indicators will serve as a criteria catalogue for evaluating applicability to horizon scanning of the proposed approach.

2.2 Existing Emotion Mining Approaches

A possible method for automated horizon scanning is to identify relevant topics by assessing their emotional mentions in Twitter micro-blogging posts. One important step for such an approach is detecting emotions in Twitter texts. This section describes existing emotion mining approaches. Emotion mining and social affective text mining deal with the detection of emotions in text. There are various different ways for doing this, such as dictionary-based and machine learning methods. Machine learning methods require a considerable amount of training data and work well with large texts [9]. However, as the aim of this work is to create an approach for short text, it focuses on dictionary based methods, specifically keyword spotting and concept-based techniques.

> **Keyword spotting** approaches rely on the mapping of single words to emotions based on emotion dictionaries. The main challenge here lies in identifying and designing the right dictionaries for the given problem context and propagating word connotations across semantically related words, such as synonyms. For example, in social media short text, dealing with social media language characteristics, such as emoticons, acronyms and misspellings is essential. Therefore, besides emotion dictionaries, emotion mining resources also include acronym dictionaries and emoticon lists. An acronym dictionary contains common social media expressions such as *idk* for *I don't know* and *omg* for *oh my god*. The mapping of emoticons to emotions, such as *{joy, :)}* can be done via emoticon lists. There are several freely available emotion dictionaries. Some of them concentrate on assigning polarity to words while others go beyond word polarity and assign emotional attributes.

> **Concept based** approaches do not focus on single words but on word concepts, which are word strings of one to several words. Semantic parsing is needed to meaningfully extract word concepts from text. These concepts can then be used to identify emotions by using a combination of dictionaries and semantic word networks for the mapping process. According to Cambria et al. [10] concept based approaches perform better than traditional dictionary-based keyword spotting methods.

Proposed emotion mining approaches commonly use a combination of natural language processing techniques, emotion dictionaries and emotion models. Natural language processing techniques, such as lemmatisation, stemming and part-of-speech tagging, are used to process given text in a way that enables further processing. Dictionaries contain emotional information for specific words or word combinations. The complexity of these dictionaries reaches from single word-to-emotion mappings, e.g. *wish – hope* to predefined lists of common negations, e.g. *not like* and adverbial phrases, e.g. *really like*. Emotion models map this emotional information to emotions, such as *fear* or *anger*. These emotion mappings can be based on models that were derived from social science studies or crowd-sourcing experiments where a number of human annotators are asked to assign emotions to language items.

The increase in availability of public unstructured data has lead to an increasing interest in analysing and understanding its contents. Existing emotion mining approaches are applied in various fields:

Reputation Management aims to measure and control the view of a brand. This can include analysis of social media, search engine results and conventional online media like newspapers and video platforms. For instance, an analysis of the first 20 entries of a search engine query could be done.

Product Review Analysis aims to analyse opinions about specific product features. For instance, a textmining program would break down a sentence such as *"The new phone has a good screen but the battery is horrible."* to *good screen* and *horrible battery*.

Online Games feature applications related to human computer interaction. For instance, changing the facial expressions of online avatars depending on their chat content.

In these fields, several ways for assigning polarity and emotional information to text content have been proposed. The following list shows some examples based on [46]:

- Discrete 2-values: Thumbs up/Thumbs down
- Discrete 3-values: Positive/neutral/negative
- Scale: 5 (strongly negative) to 5 (strongly positive)
- Dual scale: 1 (no positivity) – 5 (strong positivity) and -1 (no negativity) - -5 (strong negativity)
- Multiple: happiness (0-100), sadness (0-100), fear (0-100).
- Discrete emotions: anger, anxiety, sadness, positive [33]

The next sections analyse state of the art emotion mining approaches.

2.2.1 EmoHeart - Conveying Emotions from Affect Sensing Text

EmoHeart [28] is an emotion approach that allows avatars in the virtual reality computer game *Second Life* to convey emotions based on their conversations' text content. It is an application of the *Affect Analysis Model* [27], which was developed to extend traditional word spotting techniques with semantic rules and phrase and sentence level analysis. The EmoHeart approach assigns emotion vectors to text content, where an emotion vector contains the strength of each discrete emotion which is part of the vector. The discrete emotions used are: anger, disgust, fear, guilt, interest, joy, sadness, shame and surprise. An example emotion vector is *e("rude")=[0.2,0.4,0,0,0,0,0,0,0];*
Figure 2.2 shows the main components of the *Affect Analysis Model* as follows:

1. Symbolic Cue Analysis
2. Syntactical Structure Analysis
3. Word-Level Analysis

4. Phrase-Level Analysis
5. Sentence-Level Analysis

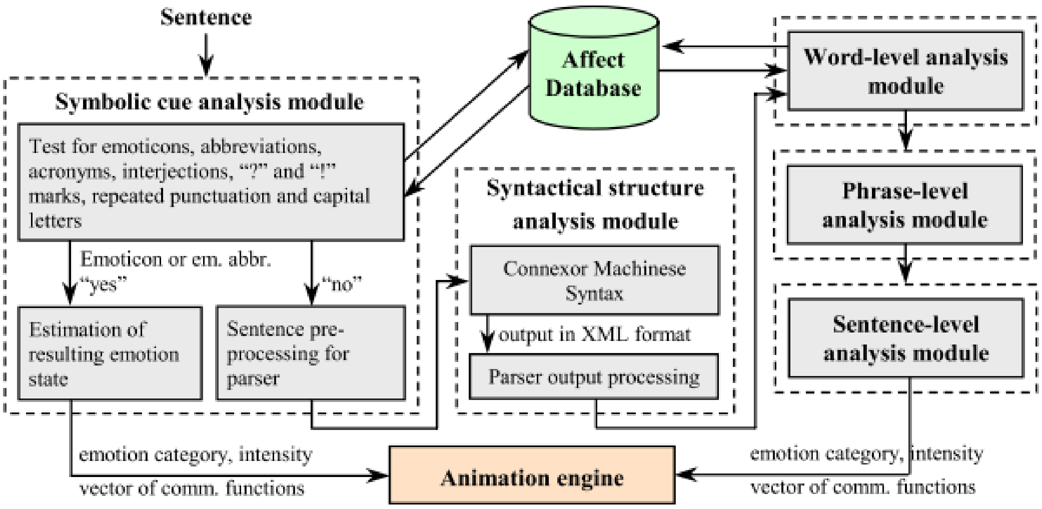


Figure 2.2: Affect Sensing Model [27]

(1) **Symbolic Cue Analysis**. During the first step of the emotion mining process, the text is scanned for symbolic cues. These include: (i) emoticons, (ii) acronyms, (iii) abbreviations, (iv) specific punctuation (repetition, exclamation mark, question mark) and (v) capital letters. The approach assumes that abbreviations and emoticons are most important in determining the affective meaning of a sentence. If they occur, all other text content is ignored. If multiple occur, the following rules apply:

(a) If they are all of the same emotional category, the intensity value for this specific emotion is increased.

(b) If the they are different, the one occurring last is given preference.

In this step, preprocessing for the next step, syntactical structure analysis, is performed. Abbreviations and acronyms are substituted with their proper transcriptions to prepare sentences for further processing. For example, *afaik* would be substituted with *as far as I know*.

(2) **Syntactical Structure Analysis**. For understanding text syntax, the Connexor Machine Syntax parser[2] is used. It takes care of identifying information such as parts of speech, syntactic function and morphological tags. Following that, sentences are described as sets of primitive clauses, such as *extremely large elephant*, *walking slowly* and *towards the jungle*.

[2]http://www.connexor.com/

(3) **Word-Level Analysis**. In this stage, an affect database with emotion dictionaries is used for mapping emotions to words. Results are stored as vectors with emotion intensity values for each of the emotions: anger, disgust, sadness, fear, guilt, interest, joy, shame and surprise. The following is an example taken from [27]:

e("rude")=[0.2,0.4,0,0,0,0,0,0,0];
e("brotherly")=[0,0,0,0,0,0,0.2,0,0];

Comparative and Superlative forms increase the emotional intensity values.

(4) **Phrase-Level Analysis**. During phrase level analysis, the text is searched the following word form co-occurrences:

- adjective phrase, e.g. *seriously worried*
- verb followed by adjective phrase., e.g. *was seriously worried*
- noun phrase, e.g. *furious anger*
- verb followed by noun phrase, e.g. *walked in the dark*
- verb followed by adverbial phrase, e.g. *(will) go in an hour*

Based on the phrase form, emotional vectors are updated according to specific rules. For example, when processing noun phrases, all emotion intensity values of each word of the phrase are compared and the maximum values are copied to a resulting phrase vector. The following example was taken from [27]:

e("love")=[0,0,0,0,0,0.8,1,0,0];
e("brotherly")=[0,0,0,0,0,0,0.2,0,0];
e("brotherly love")= =[0,0,0,0,0,0.8,1,0,0]);

Respectively, there are rules for the following modifiers:

- adverbs of degree, e.g. *very*
- negation, e.g. *not*
- preposition, e.g. *except*

For example, the occurrence of negating modifiers leads to neutralisation of the emotion vector. For instance,

e("rude")=[0.2,0.4,0,0,0,0,0,0,0]
e("not rude")=[0,0,0,0,0,0,0,0,0]

(5) **Sentence-Level Analysis**. In this final stage, an emotional vector is generated for each sentence. A sentence has a subject and one to many verb-object formations. The rules applied here are similar to those on phrase-level. For example, if emotional vectors for subject and verb-object are opposite, the verb-object phrase is given preference and its emotional vector is considered. For instance,

My darling (Subject) smashed (Verb), his guitar (Object) [27]

will lead to a unification of the emotion vectors of *smashed* and *guitar* which will be stored as the resulting sentence vector.

EmoHeart was evaluated by using data as gold standard where two of three annotators agreed and calculating precision and recall. The results are shown in table 2.2.1.

Measure	joy	sadness	anger	disgust	surprise	fear	neutral
Precision	0.846	0.673	0.910	0.946	0.758	0.785	0.698
Recall	0.858	0.763	0.564	0.506	0.652	0.730	0.862

Table 2.2: Affect Analysis Model Evaluation [28]

2.2.1.1 EmoHeart Summary

Technique	Result Type	Resources	Emotion Model	Corpus
Word spotting	Emotions and Valence	1627 words (WNA), potential emotion eliciting words, interjections, 364 emoticons, 337 acronyms, 122 modifiers	Izard 9 emotional states: anger, disgust, fear, guilt, interest, joy, sadness, shame, surprise	160 sentences from online diary-like blog posts

Table 2.3: EmoHeart Features

2.2.2 Emotion Mining for Facebook Friendship Classification

The friendship classification framework by Mohamed and Hazem [49] uses lexical features to train a machine learning classifier. Despite the fact that it proposes a machine learning approach rather than working with keyword spotting, its use of lexical features makes it relevant to this thesis. The paper states common challenges in emotion mining of social media short texts and proposes a framework for classifying friendship relations on Facebook. The identified challenges are:

- Intentional misspelling, e.g. *hellooooo*
- Interjections and lexical surrogates, e.g. *hmmm*
- Use of upper-case letters, e.g. *COOL*
- Excessive use of punctuation, e.g. *cool!!!!*

- Social acronym, e.g. *brb* for *be right back*

- Emoticons, e.g. *:)* for indicating joy

Following that, a framework which is depicted in figure 2.3 was proposed. Based on the challenges above, the framework uses certain lexical features to assess text subjectivity. The lexical features are: repeated letters, number of emoticons, rating of emoticons, number of acronyms, number of affective words and rating of affective words.

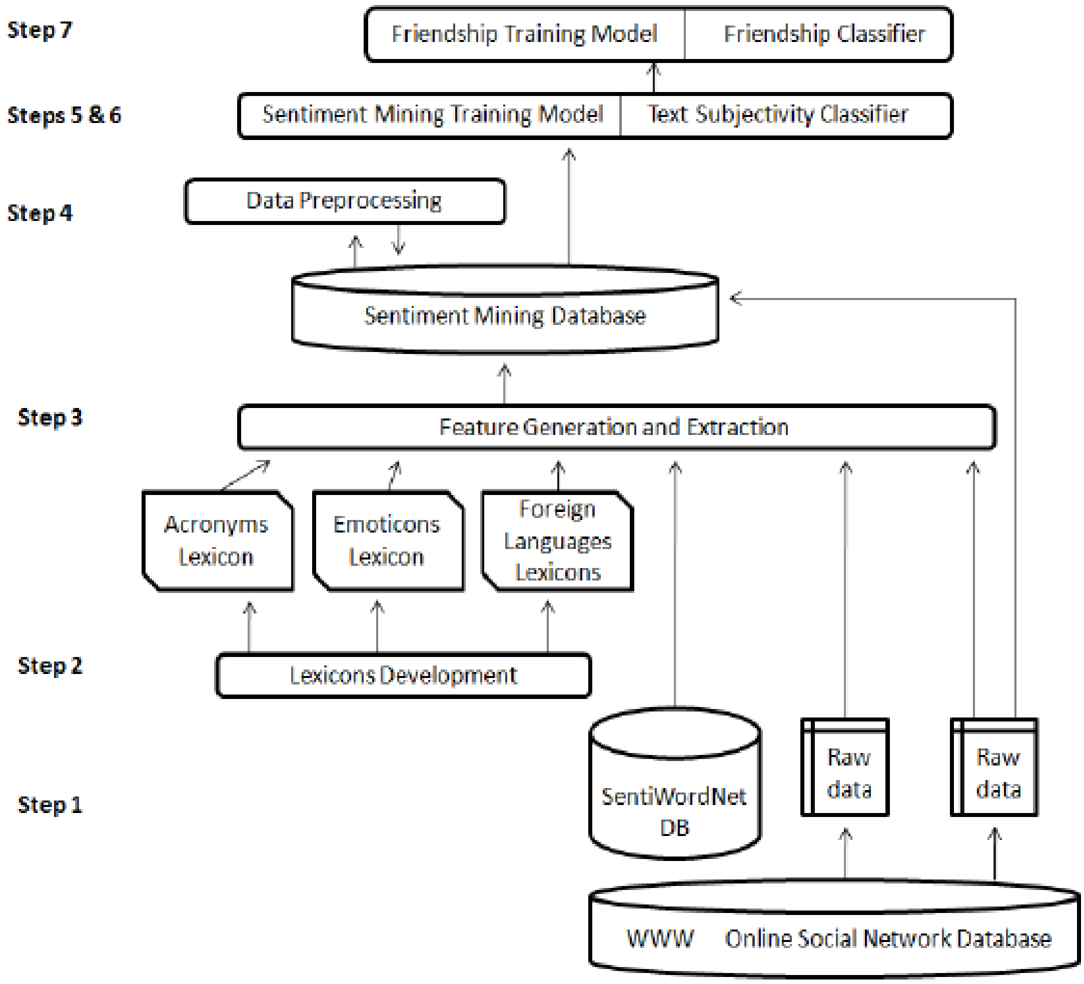


Figure 2.3: Framework for Friendship Classification [49]

The framework uses SentiWordNet for calculating the number of affective words in Facebook comments and rating them on an emotion scale. Further details on SentiWordNet are described in section 2.3.2. For acronyms and emoticons, the authors have developed their own lexicons. Both of them are mapped to terms, which in turn are mapped to emotions through the use of SentiWordNet.

2.2.3 SentiStrength - Sentiment Strength Detection

SentiStrength [46] is a sentiment mining approach that uses a dual scale, 1 (no positivity) – 5 (strong positivity) and -1 (no negativity) – -5 (strong negativity). Its aim is not to classify text by its associated emotions but to use polarity values for assigning positivity and negativity. The authors chose this system based on a study by Norman et al. [29], which proposes that people can experience positive and negative emotions simultaneously. The reason this system is examined in this work even though it does not work with traditional emotions, such as fear or anger, is that it uses a number of linguistic rules for improving its polarity scoring. Thus, it goes beyond only using word dictionaries.

SentiStrength uses a dictionary consisting of words from the lexical resources LIWC [45] and General Inquirer [41]. In order to simplify preprocessing steps before looking up words in the dictionary, the dictionary contains word stems with kleene stars as wildcards. For example, *amaz** matches *amazing*, *amazed*, *amazon*, etc. When processing text, SentiStrength first extracts single words, emoticons and punctuation and then uses the word dictionary for looking up polarity values. The overall score is not the total sum of all the words in a sentence, but simply the lowest negative (-5) and the highest positve (+5) polarity score of words in this sentence. The following illustrates this in an example:

> *"Mike is horrible and nasty but I am lovely. I am fantastic."* would be classified as *Mike is horrible[-4] and nasty[-3] but I am lovely[2]* with a sentence score of *<2,4>* and *I am fantastic[3]* with a sentence score of *<3,-1>*. The numbers in square brackets indicate sentiment strength of the preceding word, and angle brackets indicate sentence scores. The overall classification for this text is the maximum positive and negative strength of each sentiment, which is *<3 ,-4>*. [46]

This process is followed by applying a few additional rule based techniques to render more precise results:

- *Idiom list.* Common Phrases are mapped by an idiom list which assigns polarity scores to them.

- *Word Disambiguation Workaround.* This approach does not intend to find the meaning of an ambiguous word, such as *like*, which may be used for expressing completely different things. While other approaches use semantic methods for word disambiguation, SentiStrength has a lexical method:

  (1) A word that can have several meanings is assigned with a neutral polarity value. At the same time, different usages of this word, which imply different meanings, are added to the idiom list. For instance, *like* has a neutral polarity score but the phrases *I like*, *you like* and *she likes* are added to the idiom list with positive scores.

(2) A word stem which functions as word stem for two entirely different words, is taken care of by double entries to the dictionary. For instance, *amaz** matches *amazon* and *amazing*. Therefore, even though it is not a word stem, amazon is added to the list and overrides false *amaz** matches.

- *Spelling Correction.* An algorithm corrects the spelling of misspelled words so they can be found in the word dictionary.

- *Repeated Letters.* If a misspelled word has repeated letters, such as in *cooool*, the polarity score gets boosted by 1.

- *Booster Word List.* Adverbial phrases are taken care of by having booster word lists containing words such as *very* that, depending on their meaning, add or subtract polarity values.

- *Negation.* Negative mentions of words are considered in separate dictionaries. For instance, *I do not hate him* does not have a negative polarity score as it is negated.

- *Emoticon List.* Emoticons are expected to express emotions and are listed and assigned with polarity scores.

- *Punctuation.* Exclamation marks boost the polarity score of the preceding phrase.

- *Consecutive negative terms.* If two negative terms are used consecutively, the overall negative polarity score is increased.

SentiStrength was evaluated by using the average annotation values of three human annotators as gold standard data and calculating correlations with computed results. The correlation coefficient for positive sentiment was $r = 0.541$ and for negative sentiment $r = 0.499$.

2.2.3.1 SentiStrength Summary

Technique	Result Type	Resources	Emotion Model	Corpus
Word spotting	Two Values: negativity and positivity	2310 words	No emotion model. Values for negativity [-5,5] and positivity [-5,5]	Twitter microblogging posts

Table 2.4: SentiStrength Features

2.2.4 UPAR7 - Emotion Mining for Headline Sentiment Tagging

UPAR7[3] [11] is an emotion mining approach which was designed during the SemEval 2007 Task 14 challenge [42]. Its aim is to assign emotions to headlines, which typically consist of a small number of words. The emotions are (1) anger, (2) disgust, (3) fear, (4) joy, (5) sadness and (6) surprise. For each news headline, an emotion vector is generated that consists of values in the interval [0,100] for each emotion. For instance, an example for such an emotion vector is: `[23,64,0,0,0,14]`. UPAR7 uses a linguistic approach, meaning it relies on lexical resources for its emotion detection. The following table shows an overview of the used resources:

- *Part-of-Speech Tagger*. The SS-Tagger [47] was used for part-of-speech tagging.
- *Semantic Parser*. The authors chose the Stanford Parser. [13]
- *WordNet* was used as semantic lexicon.
- *WordNet-Affect* was used for assigning emotional labels to terms.
- *SentiWordNet* was used for assigning valence to terms.

UPAR7 uses three main steps for emotion mining:

(1) **De-capitalization** is used as a necessary step for preprocessing in order to improve of further semantic text parsing. This is done by using the results from part-of-speech tagging, semantic information from WordNet and some additional rules. An example rule is that a word with no existing WordNet lemma is considered as a proper noun and keeps its capitalised letter. The tools used for this step are the Stanford Parser [13], as it was tested to be more tolerant with grammatically incorrect sentences than the Link Grammar Parser [19].

(2) **Rating of individual words** is done by looking up individual terms in WordNet-Affect and SentiWordNet and boosting specific emotions. In order to extract the individual lemmas from the given text, morphology functions of WordNet are used. Following that, instead of disambiguating word lemmas, the authors chose to compute the linear combination of all possible meanings of a word balanced by the frequency of its specific lemmas. The boosting of specific emotions is done by applying specific linguistic rules, such as assuming that the occurrence of question marks and modal auxiliaries is an indicator for surprise. Another example is using hyponym relations to boost specific emotions. For instance, the emotions fear and and sadness are strengthened in case the given noun is somewhere in the hyponym chain of the noun *unhealthiness*.

(3) **Rating of sentences** is done through additionally boosting found emotions from step (2) by finding the main subject of the given text. Semantic dependency graphs were used to identify the word that is never a dependent word, which is considered the main word of the clause. Following that, the emotion and valence values of this main word are multiplied by 6.

[3] UPAR7 stands for University Paris 7

UPAR7 was evaluated by taking average annotation values of 6 annotators and calculating correlations, precision and recall. The results are shown in table 2.2.4.

	Pearson	Precision	Recall
Anger	32.33	16.67	1.66
Disgust	12.85	0.00	0.00
Fear	44.92	33.33	2.54
Joy	22.49	54.54	6.66
Sadness	40.98	48.97	22.02
Surprise	16.71	12.12	1.25

Table 2.5: UPAR 7 Evaluation [11]

2.2.4.1 UPAR7 Summary

Technique	Result Type	Resources	Emotion Model	Corpus
Word spotting	Discrete Emotions and Valence	1627 words (WNA), 568 emotional words , acronym list	6 emotions: anger, disgust, fear, joy, sadness, surprise	1000 news headlines

Table 2.6: UPAR 7 Summary

2.2.5 SenticNet Applications

SenticNet [10] is an opinion mining and sentiment analysis resource that was developed to produce more accurate results than plain keyword spotting techniques [10]. Unlike the aforementioned techniques, applications of SenticNet use a dictionary with word combinations, also referred to as *word concepts*, rather than single words. This dictionary is called SenticNet, which was later extended to SenticNet 2 [6] and can be used for emotion mining. However, its authors also proposed an approach of how this resource can be used in various contexts. Opinion mining using SenticNet is made up of the following components [5]:

- *NLP Module.* In the first step, natural language processing is done in order to extract word lemmas.

- *Semantic Parser*. A semantic parser parses the lemmatised text and outputs a list of word concepts, which can consist of one to three words.

- *SenticNet.* SenticNet is semantic lexical database containing word concepts and affective labels with intensity values. It is further described in section 2.3.5.

- *Hourglass of Emotions.* The Hourglass of Emotions was used to map the affective labels to a predefined set of emotions.

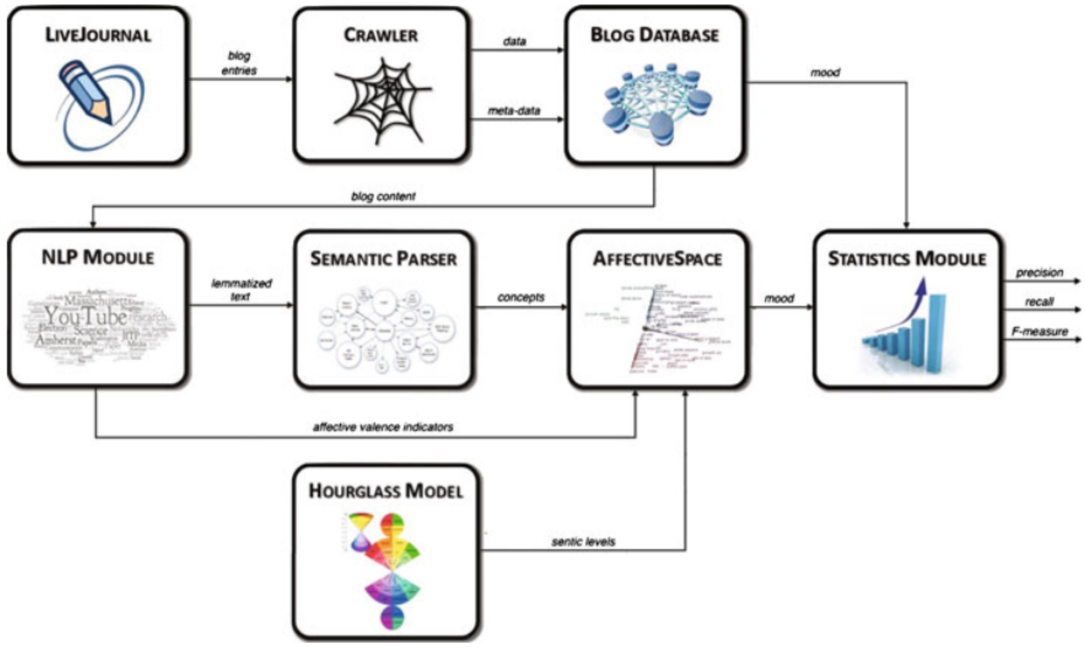


Figure 2.4: SenticNet

(1) **Natural Language Parsing**. In the first step of the sentics extraction process, the text is processed according to specific features. These include: (i) specific puncuation, (ii) complete upper-case words, (iii) repetitions, (iv) negation, (v) degree adverbs and (vi) emoticons. Following this preprocessing, the text is lemmatised and passed to the semantic parser.

(2) **Semantic Parsing**. In this step, the lemmatised text is broken down to word combinations. In SenticNet, word combinations are called word concepts and several concept extraction approaches have been proposed, for instance:

   (a) Concept extraction using a graph-based approach [38]

   (b) Concept extraction using dependency-based semantic parsing [36]

   An example for concept extraction is transforming a sentence, e.g. *I was so tired in the morning because I do not like Mondays* to (i) *so tired*, (ii) *morning* and (iii) *not like Monday*.

(3) **Concept Rating**. Finally, each extracted concept is looked up in the SenticNet dictionary, which lists pleasantness, attention, sensitivity and aptitude for each concept. An extensive explanation of the SenticNet dictionary can be found in section 2.3.5. Average values of each dimension for all concepts in the whole text are computed to give an overall emotion vector for a text. This is also referred to as sentic vector.

```xml
<rdf:Description rdf:about="http://sentic.net/api/en/concept/a_lot_of_fat">
    <rdf:type rdf:resource="http://sentic.net/api/concept"/>
    <text xmlns="http://sentic.net/api">a lot of fat</text>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/tasty"/>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/flavor"/>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/sweetness"/>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/edible"/>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/white_brown"/>
    <pleasantness xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.038</pleasantness>
    <attention xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0</attention>
    <sensitivity xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.05</sensitivity>
    <aptitude xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.092</aptitude>
    <polarity xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.027</polarity>
</rdf:Description>
```


Figure 2.5: Sample SenticNet Entry

There is no conclusive peer-reviewed evaluation of applications of SenticNet which classify text by emotions.

2.2.5.1 SenticNet Summary

Technique	Result Type	Resources	Emotion Model	Corpus
Word concept spotting	5 dimensions [-1,1]	13743 word concepts (SenticNet)	Hourglass of Emotions: pleasantness, attention, sensitivity, aptitude, polarity	5000 LiveJournal blog posts (150 words each)

Table 2.7: SenticNet Features

2.2.6 Comparison and Overview

Approach	Technique	Coverage	Emotion Model	Corpus
Affect Analysis Model	word spotting, phrase-level analysis	1627 words (WNA), potential emotion eliciting words, interjections, 364 emoticons, 337 acronyms, 122 modifiers	Izard 9 emotional states: anger, disgust, fear, guilt, interest, joy, sadness, shame, surprise	160 sentences from online diary-like blog posts
UPAR7	word spotting	1627 words (WNA), 568 emotional words , acronyms	6 emotions: anger, disgust, fear, joy, sadness, surprise	1000 news headlines
SentiStrength	word spotting	2310 words	Negative and Positive (-5 to 5)	Twitter microblogging posts
SenticNet	word concept spotting	13743 word concepts	Hourglass Model 4 dimensions -> 24 emotions	LiveJournal blog posts

Table 2.8: Comparison of Emotion Mining Approaches

Comparison of Emotion Mining Approaches. Emotion Mining for Facebook Friendship Classification is not included as it uses machine learning.

2.3 Emotion Mining Resources

Dictionary based emotion mining approaches rely on lexical resources for assigning emotions to words, phrases and sentences. The research community has proposed several of these dictionaries. They can be differentiated by size, used emotion models, the presence of a valence dimension and public availability. Furthermore, they may feature specific information that helps with common problems such as word ambiguity when a single word has two or more different meanings.

2.3.1 WordNet Affect

WordNet Affect [43] is a lexical resource which assigns affective labels to WordNet synsets. Its an extension of *WordNet Domains*, a system for labelling WordNet synsets with specific hierarchy labels. An example for a label hierarchy in WordNet Domains is *Economy ->Finance ->Banking*. The WordNet synset *bond#2* would then be labeled with the *Banking* domain label. For WordNet Affect (WNA), a similar hierarchy was developed for affective terms and an example is illustrated in 2.9. Furthermore, the authors differentiated between categories of affect as depicted in table 2.10.

emotion			
	positive emotion		
		joy	
			amusement

Table 2.9: Example of a-label hierarchy for affective category *emotion*

Category	# Example Term
emotion	anger
cognitive state	doubt
trait	competitive
behaviour	cry
attitude	skepticism
feeling	pleasure

Table 2.10: WordNet Affect Categories [43]

For applying the affective labels to WordNet synsets, the authors composed a lexical dictionary called *Affect* containing 1903 terms. For each synset which contains at least one of these words, the affective labels were applied accordingly. Furthermore, semantic WordNet relations were examined and affective labels were propagated along these relations, if applicable. It was found that the relations where affective meaning was preserved were the following: *antonymy, similarity, derived-from, pertains-to, attribute* and *also-see*. The final version of WNA contains 2,874 synsets and 4,787 words but only a subset of 627 synsets and 1627 words is labelled as emotions and can be found in the WNA emotion lists:

Category	Example	# of Synsets
anger	frustration	129
disgust	hideous	20
fear	shy	86
joy	cheerfully	235
sadness	sorrow	127
surprise	wonder	31

Table 2.11: WordNet Affect Coverage: 1627 words, 627 synsets

2.3.2 SentiWordNet

SentiWordNet is a lexical dictionary based on WordNet which was created with machine learning methods. Therefore, its coverage is very high, as it contains all WordNet synsets. The first version was based on WordNet 2.0 while the most current version is based on WordNet 3.0, containing 155,287 words and 117,659 synsets. SentiWordNet was developed for sentiment analysis and assigns two scores for each synset:

PosScore, indicating the degree of positive connotation
NegScore, indicating the degree of negative connotation
ObjScore, indicating the degree of objective connotation

The sum of PosScore, NegScore and ObjScore is always 1.

Category	# Words
PosScore	155287 words and 117659 synsets
NegScore	
ObjScore	

Table 2.12: SentiWordNet Statistics

2.3.3 GI - The General Inquirer

The General Inquirer was first developed in 1966 [41] and is a lexical resource containing 13000 words and 34914 word senses [30]. The words are assigned to specific labels within certain ontologies. The *tag* ontology groups emotional words into positive and negative terms while the *marker* ontology assigns them to 182 syntactic categories. The *tag* and the *marker* ontologies are not mutually exclusive. Marker ontology categories related to emotion mining are *Emotions* with labels such as *anger, fury, happy* and *Evaluate Adjectives* with labels such as *good, bad, beautiful*.

2.3.4 LIWC - Linguistic Inquiry and Word Count

LIWC is a a software package that was originally developed for an exploratory study of language and led to the development of the LIWC dictionary as lexical resource. »>HERE REFERENCE

Category	Examples	# Words
Emotion	Adore, Admiration, Affection	311
Evaluative Adjective	Abominable, Admirable, Adorable	205

Table 2.13: General Inquirer Categories

OF PEOPLE USING IT««. It was expanded in LIWC2001 [34] and later further extended in LIWC2007 [33], now containing about 4500 word stems, which are assigned to over 50 word categories. The word categories are hierarchically organised and contain grammatical as well as contextual information. Examples for language categories are: *Articles, Prepositions, Negations, Conjunctions* and *Numbers*. Other categories include: *Psychological Processes, Perceptual Processes* and *Biological Processes*. Categories that may be relevant to emotion mining can be found under *Psychological Processes*, which has the subcategory *Affective Processes* with the subcategories *Positive emotion* and *Negative Emotion*. Negative emotions are further split into *Anxiety*, *Anger* and *Sadness*. In total, the *Affective Processes* label applies to 915 words. See the distribution of affective word labels in the following table:

Category	Examples	# Words
Affective processes	Happy, cried, abandon	915
Positive emotion	Love, nice, sweet	406
Negative emotion	Hurt, ugly, nasty	499
Anxiety	Worried, fearful, nervous	91
Anger	Hate, kill, annoyed	184
Sadness	Crying, grief, sad	101

Table 2.14: LIWC Emotion Dictionary [33]

The LIWC dictionary is part of the LIWC software package and is not freely available.

2.3.5 SenticNet

The semantic lexical database SenticNet was developed along with an approach for mining emotions in text [10]. Its aim is to enable going beyond word-level natural language processing by considering related opinions based on common sense knowledge. Thus, it is made up of common sense knowledge concepts such as *a lot of fat*, *bad time*, *make parent happy*, *traffic jam* and *zone out*.
A typical entry of the SenticNet database has:

(a) a text attribute containing the word concept: *a lot of fat*

(b) semantically and affectively related concepts: *tasty*, *flavor*, *edible*

(c) numerical values in the range of -1 to 1 for pleasantness, attention, sensitivity, aptitude and polarity. An example SenticNet entry is shown in figure 2.6.

```xml
<rdf:Description rdf:about="http://sentic.net/api/en/concept/bad_time">
    <rdf:type rdf:resource="http://sentic.net/api/concept"/>
    <text xmlns="http://sentic.net/api">bad time</text>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/deliberate"/>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/unwanted_outcome"/>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/say_eureka"/>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/look_fact"/>
    <semantics xmlns="http://sentic.net/api" rdf:resource="http://sentic.net/api/en/concept/rational"/>
    <pleasantness xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">-0.392</pleasantness>
    <attention xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0</attention>
    <sensitivity xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">-0.327</sensitivity>
    <aptitude xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0</aptitude>
    <polarity xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">-0.24</polarity>
</rdf:Description>
```


Figure 2.6: SenticNet Entry for *bad time*

The SenticNet database contains **13743 word concepts**.

The authors propose using Plutchik's emotion model [35] for assigning these numerical values to emotions. Plutchik's model describes eight basic emotions, each one having an opposite emotion: *ecstasy-grief*, *vigilance-amazement*, *rage-terror* and *loathing-admiration*. For these emotions, their are 3 levels of intensity, for example *pensiveness* can become *sadness* and *sadness* can turn into*grief*. The emotions are organised in a wheel, so that each one has two neighbours. Combined with a neighbour, an emotion can build a new emotion. For example, *sadness* and *surprise* lead to *disapproval* and *joy* and *trust* lead to *love*.

SenticNet uses and adaption of Plutchik's model and organises the emotional state along four independent dimensions [8]:

1. Pleasantness: relating to a person's happiness

2. Attention: relating to a person's interest

3. Sensitivity: relating to a person's is comfort

4. Aptitude: relating to a person's confidence

These four dimensions can be further processed in various ways:

- **Polarity Calculation** Calculation of a *polarity* value:
  $p = \sum_{i=1}^{N} \frac{Pleasantness(c_i)+|Attention(c_i)|-|Sensitivity(c_i)|+Aptitude(c_i)}{3N}$

- **Emotion Mapping** Mapping to 24 emotional states of mind by using table 2.15.

- **Second Level Emotion Mapping**. Deriving second level emotions which are compound emotions depending on two Sentic dimensions as listed in table 2.16:

The original wheel-shaped model of Plutchik becomes hourglass-shaped as depicted in figures 2.7 and 2.8.

Interval	Pleasantness	Attention	Sensitivity	Aptitude
$[\frac{2}{3}, 1]$	ecstasy	vigilance	range	admiration
$[\frac{1}{3}, \frac{2}{3}]$	joy	anticipation	anger	trust
$[0, \frac{1}{3}]$	serenity	interest	annoyance	acceptance
$[-\frac{1}{3}, 0]$	pensiveness	distraction	apprehension	boredom
$[-\frac{2}{3}, -\frac{1}{3}]$	sadness	surprise	fear	disgust
$[-1, -\frac{2}{3}]$	grief	amazement	terror	loathing

Table 2.15: SenticNet Dimension-Emotion Mappings [7]

	Attention>0	Attention<0	Aptitude>0	Aptitude<0
Pleasantness>0	optimism	frivolity	love	gloat
Pleasantness<0	frustration	disapproval	envy	remorse
Sensitivity>0	aggressiveness	rejection	rivalry	contempt
Sensitivity<0	anxiety	awe	submission	coercion

Table 2.16: Table taken from [7]

2.3.6 EmoSenticNet

EmoSenticNet [37] is an emotion dictionary which assigns Ekman emotions [15] to SenticNet concepts. As described in section 2.3.5, entries of the SenticNet database are mapped to emotion dimensions rather than discrete emotions. The Hourglass Model of Emotions [8] proposes ways of translating these dimensions to emotions such as *interest* and *distraction*. However, the proposed techniques contain 16 and 24 emotions, leading to a large number of different emotions in the result set of an emotion mining approach. Alternatively, EmoSenticNet uses a machine learning approach which combines information from SenticNet (section 2.3.5), WordNet Affect (section 2.3.1) and ISEAR[4] to assign six discrete emotions to SenticNet concepts. The emotions are: *anger, disgust, joy, sadness, surprise* and *fear*.

2.3.7 SentiSense

SentiSense [12] is an emotion dictionary that aims to address the shortcomings of other emotion dictionaries, such as LIWC [45], LEW [18] and WordNet Affect [43]. All of them assign emotional values to single words and therefore have several disadvantages.

Word Ambiguity

First, word disambiguation is a problem. There are words which may be used for several meanings, for example, the adjective: *cool*. The WordNet dictionary [17] proposes several definitions for the term *cool* as shown in figure 2.3.7. Two Examples are:

(a) neither warm nor very cold; giving relief from heat

(b) being satisfactory or in satisfactory condition

[4]ISEAR: International Survey of Emotion Antecedents and Reactions

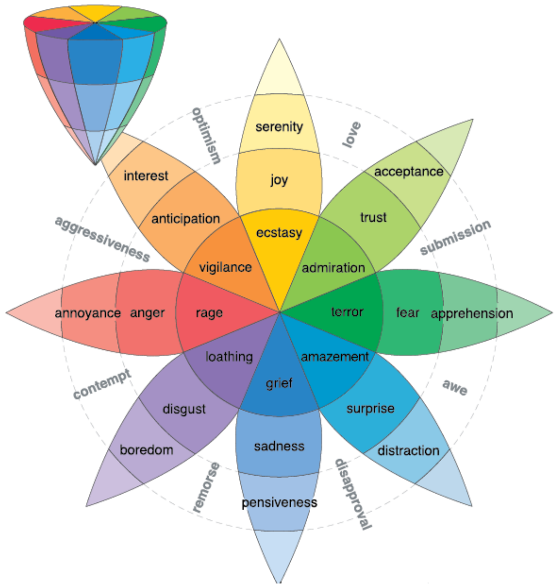


Figure 2.7: Wheel of Emotions

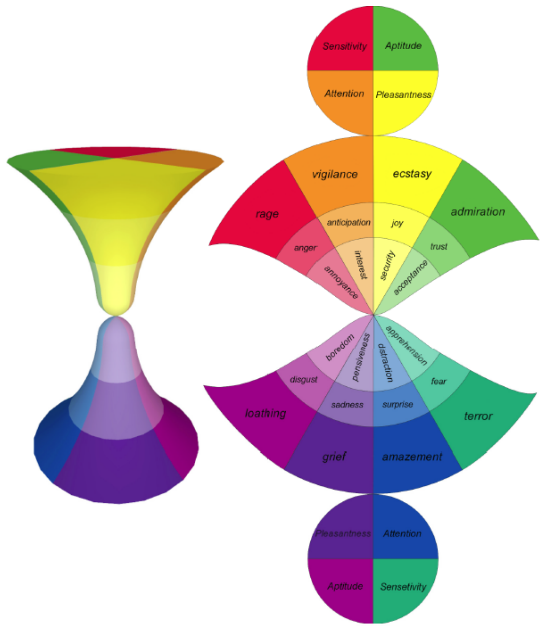


Figure 2.8: Hourglass of Emotions

While someone may assign a rather neutral emotion to the meaning of (a), the meaning of (b) may be considered as something associated with a positive emotion. WordNet Affect [43] does not address this problem.

- S: (adj) **cool** (neither warm nor very cold; giving relief from heat) *"a cool autumn day"; "a cool room"; "cool summer dresses"; "cool drinks"; "a cool breeze"*
- S: (adj) **cool**, coolheaded, nerveless (marked by calm self-control (especially in trying circumstances); unemotional) *"play it cool"; "keep cool"; "stayed coolheaded in the crisis"; "the most nerveless winner in the history of the tournament"*
- S: (adj) **cool** (inducing the impression of coolness; used especially of greens and blues and violets when referring to color) *"cool greens and blues and violets"; "the cool sound of rushing water"*
- S: (adj) **cool** (psychologically cool and unenthusiastic; unfriendly or unresponsive or showing dislike) *"relations were cool and polite"; "a cool reception"; "cool to the idea of higher taxes"*

Figure 2.9: WordNet Definition for Adjective cool

Non-exclusiveness

Furthermore, in 113 out of 911 synsets [12], several emotions have been assigned to them. Therefore, Non-exclusiveness can be identified as another problem that complicates interpretation of computed emotion mining results. The authors of the SentiSense approach tried to overcome this limitation by assigning emotions to WordNet synsets rather than single terms and using word sense disambiguation algorithms to detect word meaning.

Vocabulary Coverage

Second, existing approaches often do not offer a high vocabulary coverage in order to keep high precision values. Approaches with high coverage, such as SentiWordNet [16], sometimes have low precision, while approaches with low coverage [18] have high precision. SentiSense aimed to achieve both high coverage and high precision by involving a large number of human annotators in the labelling process of emotional words. Table 2.17 lists the numbers of synsets per emotion in SentiSense.

Category Example	# Synsets
like	426
joy	164
anger	64
hate	18
love	68
despair	13
anticipation	165
sadness	174
calmness	69
ambiguous	60
fear	199
hope	68
surprise	40
disgust	662

Table 2.17: 5496 words and 2190 synsets

Category	Antonym	Category	Antonym
Ambiguous	-	Hate	Love
Anger	Calmness	Hope	Despair
Calmness	-	Joy	Sadness
Despair	Hope	Like	Disgust
Disgust	Like	Love	Hate
Anticipation	Surprise	Sadness	Joy
Fear	Calmness	Surprise	Anticipation

Table 2.18: SentiSense Antonyms. Table taken from [12]

The emotion model used by SentiSense resulted from a combination of three models proposed by Arnold, Plutchik and Parrot and further editing by computational linguists who made sure selected emotion labels have clear antonyms with the list of emotions. Furthermore, they removed emotions such as *dejection* that were not expected to be commonly used in opinionated texts. The resulting list of emotions and their corresponding antonyms is described in table 2.18.

The authors examined semantic relationships between words and concluded that emotional labels could be propagated along the following semantic relations: derived-from-adjective, pertains-to-noun-participle-of-verb. They also assigned opposite emotion labels to antonyms.

2.3.8 Comparison and Overview

Table 2.3.8 features a comparison of affect dictionaries.

Dictionary	Size	Emotions	Valence	Available?
WordNet Affect	1627 words, 627 synsets	6 emotions: anger, disgust, fear, joy, sadness, surprise	no	yes
SentiWordNet	155287 words, 117659 synsets	3 variables: positive, negative, objective	yes	yes
DAL	8742 words	3 dimensions: pleasantness, activation, imagery	no*	no
GI	516 words	2 categories: emotion and evaluative adjective	no	no
LIWC	915 words	4 emotions: positive, anxiety, anger, sadness	no	yes
SentiSense	5496 words, 2190 synsets	14 emotions + antonyms	no	yes
SenticNet	13743 word concepts	4 dimensions that can be translated to 16 or 24 emotions	no*	yes

Table 2.19: Comparison of Lexical Resources

* Dimensional approaches offer valence of their proposed dimensions. They propose methods to translate dimensions to emotions but without clues of how to retain valence.

CHAPTER 3

Design

3.1 Design Considerations

The aim of this work is to create an emotion mining approach which (1) uses emotion dictionaries, (2) deals with requirements for Twitter micro-blogging posts and (3) outputs the results in a format which is usable for the horizon scanning process. Based on these three main requirements a literature survey was performed in chapter 2.1.

The emotion mining approach itself (1), required the most effort as current research does not propose a universally accepted method for emotion mining. Instead, available work uses a wide range of different lexical resources, natural language processing and dependency parsing techniques. Many of these resources are not publicly available and papers lack exhaustive information about employed custom dictionaries. Furthermore, they were applied to a variety of different text forms, such as newspaper headlines and large blog entries. Different text forms may require different approaches and further contribute to the complexity of designing am emotion mining approach.

Moreover, as an important aim was to generate a usable result, technical feasibility was a major criterion. Therefore, the following factors had to be fulfilled:

(a) **Availability of Lexical Resources**. The used emotion dictionary has to be publicly available and accessible via common data interfaces. Once availability of dictionaries was ensured, it had to be decided how to work with them. Options considered were:

(1) Focus on a small subset of emotions and merge dictionaries with respect to these emotions

(2) Use different dictionaries for different purposes

(3) Custom dictionaries. Include separate dictionaries for interjections, emoticons, swear words and social acronyms.

(b) **Acceptable Performance**. The emotion rating process should not take longer than 250 milliseconds per Tweet.

(c) **Suitable Emotion Model**. The used emotion model has to make sense with regard to the desired application to support the horizon scanning process. Questions asked where: What emotion model does the emotion dictionary build on? Can it be applied to other emotion models? How? What rating strategies could be used with this emotion model? Could this emotion model be interesting for horizon scanning purposes?

(d) **Rich Meta-Data**. As described in section 2.1, showing the sole occurrence of emotional content is not enough to support horizon scanning. Other meta-data, such as location and time play a vital role in analysing the results of the emotion detection mechanism. A researcher would cluster the emotion mining results according to different meta-data dimensions before making decisions about potential weak signals. For example, the changes of emotions associated with the most popular hashtags in the UK between November and March could be examined on an area chart as shown in section 5.3. An analysis of section 2.1 results in a series of relevant meta-data: topic, emotion, location, time, person, organisation.

The following is a short analysis of how the different types of meta-data which are relevant to horizon scanning are collected by the proposed approach.

Topic information can be extracted using Twitter hashtags.

Emotion information can be generated by the designed emotion mining approach.

Location information can be taken from Twitter source data.

Time information can be taken from Twitter source data.

Person is neglected as Twitter information about post authors is ambiguous.

Organisation is neglected as there is no simple way of tracing Tweet origins to organisations.

(e) **Leveraging Twitter Data Characteristics**. As the emotion mining approach is optimised for micro-blogging posts (Tweets) taken from Twitter, the following characteristics[1] should be considered:

- **Tweet Length** is limited to 140 characters
- The at sign followed by a username transforms it to a link to the user's profile
- **#** The hash symbol is used to assign a topic to a Tweet. Twitter users can search for so called hashtags.
- **RT** Indicates that the user's Twitter post is reTweeted.
- **Short URLs** are used to shorten URLs in Tweets .

[1] https://support.twitter.com/articles/166337-the-twitter-glossary

- **Images** can be uploaded as part of a Tweet. They are added to the Tweet in URL format (pic.twitter.com + <char code>) and form part of the total character count.
- **Grammar mistakes** may occur, as people might try to express an opinion quickly or they want to wrap as much information in one Tweet as possible.
- **Acronyms** are commonly used in social media language.
- **Emoticons** are often used to indicate emotion and sarcasm.

3.2 Choosing Lexical Resources and Techniques from Existing Approaches

Decisions regarding used emotion mining resources and techniques were made based on the considerations described in section 3.1. This section briefly describes which design elements were chosen and the reasons why specific design decisions were made. An exhaustive explanation of all architecture components can be found in section 3.3.

Emotion Dictionaries are lexical resources used for assigning emotions to words. SenticNet is used as main emotion dictionary as it (1) is still being extended, (2) has a large number of entries, and (3) contains word combinations rather than single words. However, as it contains four dimensions rather than discrete emotions, EmoSenticNet was used for mapping SenticNet entries to discrete Ekman emotions. SentiSense was used to cross-check emotion labels as it deals with the problem of word ambiguity.

Weighting Strategies are used to descide how to combine emotional values of words across phrases, sentences and the whole Tweet. Weighting is done similar to SentiStrength, were the strongest emotion is used for the overall emotion of a text piece. Furthermore, some rules for boosting emotions were applied, which is also done in EmoHeart and UPAR7. Moreover, similarly to EmoHeart, the proposed approach adopts an initial analysis for symbolic cues, such as emoticons, and starts the emotion-dictionary-based process only if no symbolic cues were found.

Custom Dictionaries are used in all analysed emotion mining methods. The proposed approach implements lists of acronyms, interjections and emoticons. In addition to that, a list of swear words is used.

Natural Language Processing also plays a crucial role for emotion mining as text needs to be preprocessed before it can be further analysed. Some techniques such as decapitalization words for better part-of-speech tagging were adopted from existing approaches. Furthermore, word tokenization was done with a Twitter-specific part-of-speech tagger.

3.3 Architecture

The proposed architecture is depicted in figure 3.3 and can be divided into 5 main parts: (1) Parsing Source Tweets, (2) Natural Language Processing, (3) Symbolic Cue Analysis, (4) Emotion Mining and (5) Visualisation.

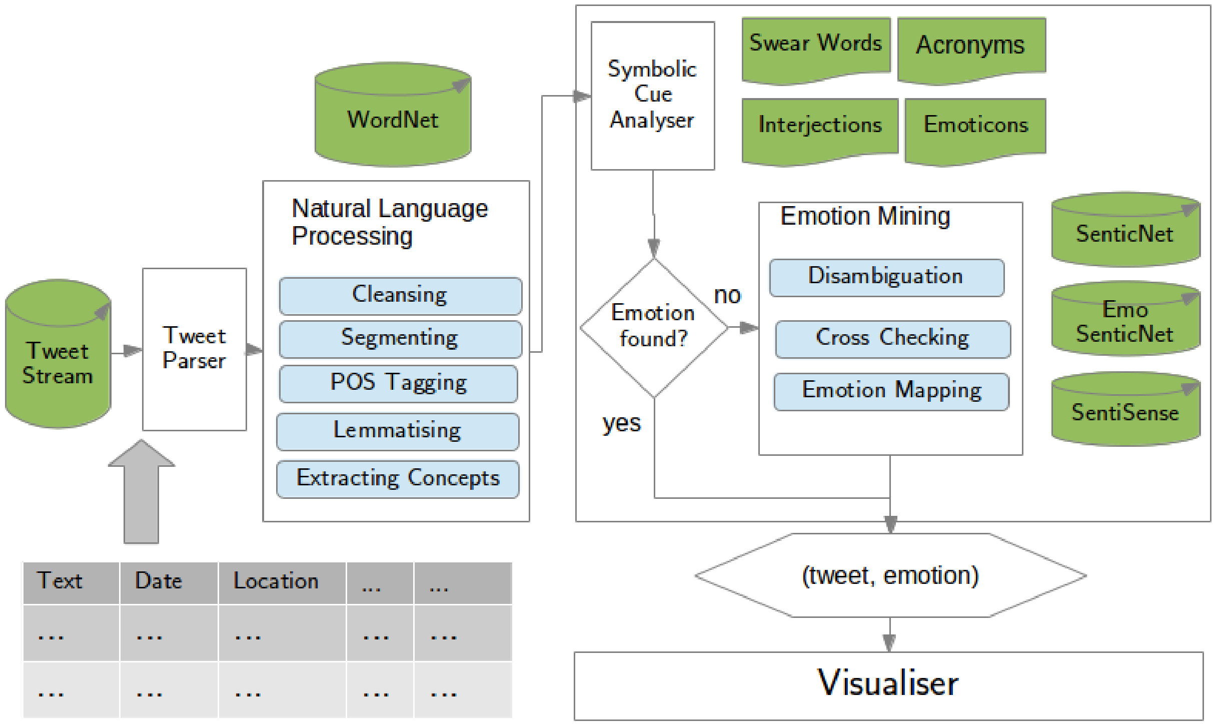


Figure 3.1: Architecture

Figure 3.2 shows the steps of the proposed emotion mining approach.

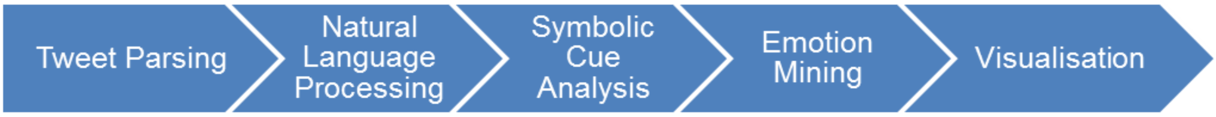


Figure 3.2: Steps of the Emotion Mining Approach

The steps of the process are as follows:

- **Parsing Source Tweets**. Tweets that need to be analysed serve as input for the program and have to be parsed. Details are described in section 3.3.1.
- **Natural Language Processing**. This module takes care of language processing activities, in particular: substitution of contractions, removal of hashtags, reduction of repeated letters, lemmatisation, part of speech tagging and concept extraction. Details are described in section 3.3.2.
- **Symbolic Cue Analysis**. During the symbolic cue analysis, custom dictionaries are used to search for strong indicators of emotion. These are: swear words, acronyms, interjections and emoticons. Details are described in section 3.3.4.
- **Emotion Mining**. If the symbolic cue analysis did not show any results, dictionaries and semantic rules are used to assign emotions to Tweets. Details are described in sections 3.3.3 and 3.3.5.
- **Visualisation**. The emotion mining results along with corresponding Tweet meta-data get visualised. Details are described in section 3.3.6.

3.3.1 Tweet Source Data

The following Tweet attributes needs to be parsed: (1) Text, (2) Location, (3) Date and Time and (4) Hashtags. Natural language processing techniques are employed to further process the input parameters.

3.3.2 Natural Language Processing

3.3.2.1 Contractions, Repeated Letters and Hashtags

Contractions, such as *don't* are expanded to their base forms, such as *do not*. There is an overall list of x contractions. Furthermore, repeated Letters are taken care of by replacing any number of a letter x with only two occurences of x. For example, *coooool* gets substitued by *cool*. Moreover, hashtags are removed.

3.3.2.2 Lemmatisation

Lemmatisation is a core feature of the designed approach. In order to enable dictionary lookups, words need to be translated to their base forms. For instance, in *I went to the markets, having a lot of fun* will be translated to *I go to the market, have a lot of fun*. This is done using the WordNet Lemmatizer, which uses information from WordNet to translate plural forms, various tense forms, gerrund forms etc. to their corresponding base forms. An alternative to lemmatisation would have been stemming, which was not used as it is subject to increasing word ambiguity. For instance, the word stem of *amazon* is *amaz*, which is the same for *amazing* and *amazed*.

3.3.2.3 Concept Extraction

The proposed approach uses SenticNet, a word concept dictionary, as its lexical resource. Therefore, a sentence has to be divided into its concepts before the dictionary can be accessed. For instance, the sentence *I prefer going on vacation over going to prison* would be split up in *go vacation* and *go prison*. The corresponding SenticNet entries for this example are:

Listing 3.1: Extract of SenticNet Entry for *go vacation*

```
<pleasantness xmlns="http :// sentic . net/api" rdf:datatype="http :// www.w3.org/2001/
    XMLSchema#float">0.92</pleasantness>
< attention xmlns="http :// sentic . net/api" rdf:datatype="http :// www.w3.org/2001/
    XMLSchema#float">0.976</attention>
< sensitivity xmlns="http :// sentic . net/api" rdf:datatype="http :// www.w3.org/2001/
    XMLSchema#float">0</sensitivity>
<aptitude xmlns="http :// sentic . net/api" rdf:datatype="http :// www.w3.org/2001/
    XMLSchema#float">0.722</aptitude>
<polarity xmlns="http :// sentic . net/api" rdf:datatype="http :// www.w3.org/2001/
    XMLSchema#float">0.873</polarity>
</rdf:Description>
```

The problem of accurately extracting concepts has been solved by the research community as described in section 2.2.5. However, since the corresponding algorithms are complex and not publicly available, a simplified version was developed for this thesis as shown in algorithm 3.1. As dictionary access performance can be maximised as described in section 4.2.2, a high accuracy for concept extraction is not necessary and can be compromised by dictionary access performance. For instance, correct concepts for the sentence *I was hungry but then decided that I need even more food later* are *be hungry* and *need food*, while the simplified algorithm extracts *be hungry*, *decide hungry*, *decide food*, *need food*.

input : A list with tuples in the format word, part of speech tag
output: A list with combinations of two words each

```
First, create lists of adjectives, adverbs, verbs and nouns;
for adjective in adjectives do
    for noun in nouns do
        Append (adjective + BLANK + noun) to list of word combinations
    end
end
for verb in verbs do
    for noun in nouns do
        Append (verb + BLANK + noun) to list of word combinations
    end
end
```

Algorithm 3.1: Code snippet showing a part of how word combinations are extracted

3.3.2.4 Twitter Part-of-Speech Tagger

As one of the aims for the proposed approach is to specialise on Twitter micro-blogging posts, a custom part-of-speech tagger is used. The CMU ARK Twitter Part-of-Speech Tagger [31] is aware of emoticons and social acronyms.

For instance, when tokenising, the widely accepted Stanford POS Tagger would transform the sentence *I am happy :)* to *I, am, happy, :,)* while the Twitter POS tagger outputs *I, am, happy, :)*.

3.3.3 Emotion Dictionaries

3.3.3.1 SenticNet

SenticNet is a lexical resource consisting of 13743 word concepts mapped to 4 dimensions: pleasantness, attention, sensitivity and aptitude (see section 2.3.5). Due to the novelty of using word concepts rather than single words thus increasing accuracy, this dictionary was used as main resource. All 4 dimensions can be calculated to an overall polarity value:
$p = \sum_{i=1}^{N} \frac{Pleasantness(c_i)+|Attention(c_i)|-|Sensitivity(c_i)|+Aptitude(c_i)}{3N}$
This polarity value is later used to assess the importance of a term as follows: A term X is deemed more emotional than term Y if its polarity p_X is higher than p_Y.

3.3.3.2 EmoSenticNet

EmoSenticNet assigns SenticNet concepts to one of the following six emotions: anger, disgust, joy, sadness, surprise, fear. The original SenticNet models features 24 emotions, leading to an emotion granularity which may overcomplicate interpretations for horizon scanning purposes. Therefore, EmoSenticNet was used for assigning emotions to word concepts, while the original SenticNet resource is merely used to assess emotional strength of a word concept as described in section 3.3.3.1.

3.3.3.3 SentiSense

SentiSense is a dictionary consisting of 5496 words within 2190 synsets mapped to 14 emotions (see section 2.3.7). Unlike the other utilised dictionaries, its entries include the respective part of speech tag. This allows to differentiate between several meanings of one word. An example is *like* which is neutral when used as and adverb (He looks like me) and positive when used as a verb (I like holidays). For this reason, SentiSense is used for word disambiguation. An emotion vector of a word is only accepted if its SentiSense entry does not suggest an emotion of the opposite polarity. A check for exact agreement of SentiSense and EmoSenticNet emotions is not performed as the dictionaries differ greatly in their assigning of emotions. While EmoSenticNet has a higher number of words labelled with *Anger* and a low number of words labelled with *Disgust*, SentiSense has features the opposite.

3.3.4 Custom Dictionaries

In a first step, symbolic cue analysis is done before a more sophisticated emotion mining approach is applied. Four different custom dictionaries were created in order to assess text subjectivity:

(1) **Acronyms**. Acronyms, such as *lol* for *laughing out loud* and wtf for for *what the fuck* were included in the dictionary and certain emotions were assigned to them.

(2) **Emoticons**. Emoticons, such as *:)* and *:(* were compiled to a dictionary. It is to be noted, that similar to words, emoticons can be grouped into synsets, for instance, *:)*, *=)*, *:-)*, *:>*, *:D*, *(:*,*(-:* and *<:* would all refer to the emotion *Joy*.

(3) **Interjections**. Interjections, such as *aargh* may also express emotions and are therefore taken care of in a separate dictionary. As they are likely to contain many repeated letters, several letter repetition combinations are considered while reading the dictionary. Examples are *argh* for *Anger* and *eeww* for *Disgust*.

(4) **Swear Words**. Although none of the analysed emotion mining approaches uses swear word dictionaries, a collection of swear words was created and assigned to emotions. Examples are *asshole* for *Anger* and *cringe* for *Disgust*

3.3.5 Weighting Strategies

Research proposes different ways for weighting emotional contents within sentences and phrases as described in section 2.2. The approach proposed in this work uses the calculated SenticNet polarity in order to assess emotional strength of a term. The term with maximum polarity is seen as indicator for the dominant emotion in a sentence. It is extracted and its corresponding emotion from EmoSenticNet 2.3.6 is assigned to the sentence. If a term does not occur in both, SenticNet and EmoSenticNet, it is not considered for emotion assessment. If a Tweet consists of several sentences, the emotion of the sentence with the strongest emotional term is used as indicator for the overall Tweet emotion. Furthermore, the following rules are applied when operating on word level:

- If the resulting emotion vector is positive but SentiSense lists the same term as a negative or ambigious emotion
  Do not consider this term for this emotion classification.

- If EmoSenticNet lists the term emotion as Disgust or Fear
  Boost the corresponding polarity value

- If EmoSenticNet lists the term as Sadness but SentiSense lists it as Disgust
  Set the term emotion to Disgust.

Moreover, the following rules are applied when operating on sentence level:

- If the strongest term has SenticNet polarity lower than a specified value (0.15)
  Classify the Tweet as `neutral`

- If the absolute term polarity of a term is the highest of all term polarities in that sentence
  Classify the Tweet as the emotion of this term

- If the the sum of all negative polarities is higher than the highest of all positive term polarities. Classify the Tweet as the emotion with the highest negative term polarity.

3.3.6 Visualisation

3.3.7 Location: Emotion Maps

A means of showing information for specific locations are geographical maps. Once Tweets have been tagged with a certain emotion, these emotions can be displayed on a map. The following image shows the use of pie charts on a map to show emotional distribution for different geographical locations. An geographical map with pie charts visualises three dimensions: location, emotions and relative quantities.

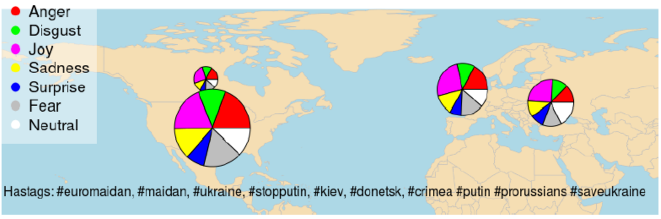


Figure 3.3: Sample Emotion Map

3.3.7.1 Time: Emotion Stacked Area Charts

For focusing on the time dimension, it is common to use line charts while using the x-axis for temporal values. To further emphasise on quantities, area charts can be used to show absolute numbers of emotional mentions of a topic. To get an overview of several emotions occuring in a similar context, stacked area charts can be used to show relative quantities as shown in figure 3.3.7.1. An stacked area chart visualises three dimensions: time, emotions and relative quantities.

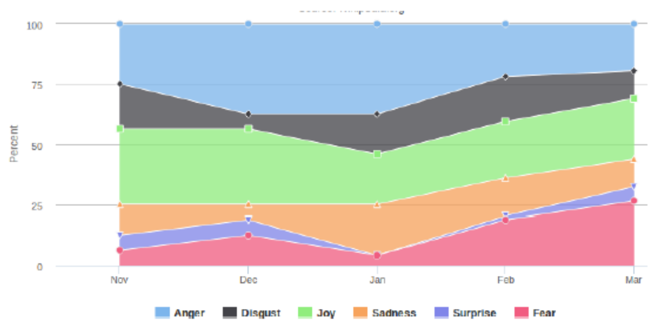


Figure 3.4: Sample Stacked Area Chart

3.3.7.2 Topic: Emotion Tree Maps and Emotion Heat Maps

If someone aims to view emotional content across different topics, tree maps can be used to show relations of different quantities of emotional mentions while also showing relations of absolute numbers of Tweets. For instance, the chart shown in figure 5.3 can be used to show the top 6 topics that are associated with the emotion *Anger*. Heat maps show relative quantities in table format, instead of visualising them in form of rectangle sizes, as its done in tree maps. Furthermore, an additional dimension can be added to a tree map by assigning meaning to different rectangle colours which results in three dimensions: colour, label and size which can be used for: topic, emotion and quantity. For instance, each rectangle in figure 5.3 for the emotion *Anger* could be further divided into smaller rectangles for different topics, represented by different rectangle colours. Heat maps offer a matrix-style listing of values, thus also supporting the display of three dimensions: row, column and value.

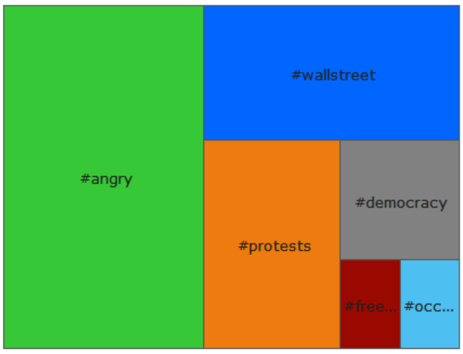
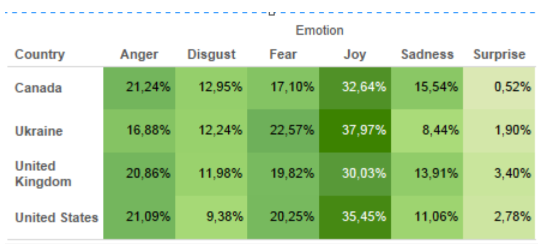


Figure 3.5: Sample TreeMap



	Emotion					
Country	Anger	Disgust	Fear	Joy	Sadness	Surprise
Canada	21,24%	12,95%	17,10%	32,64%	15,54%	0,52%
Ukraine	16,88%	12,24%	22,57%	37,97%	8,44%	1,90%
United Kingdom	20,86%	11,98%	19,82%	30,03%	13,91%	3,40%
United States	21,09%	9,38%	20,25%	35,45%	11,06%	2,78%

Figure 3.6: Sample HeatMap

CHAPTER 4

Prototype Implementation

The proposed emotion mining design of section 3 was used to develop a prototype. This section describes technical details.

4.1 Software Environment

The following software was installed

Operating System	Ubuntu 12.04.4 LTS
Programming Language	Python 2.7.3
Natural Language Processing	Python NLTK 2.0.4
POS Tagger and Tokenizer	Twitter POS ark nlp 0.3.2 Stanford Parser 3.1.5
Converting RDF to NTriples and dealing reading NTriples format	redland-1.0.17 rasqal-0.9.32 raptor2-2.0.14.tar.gz redland-bindings-1.0.17.1.tar.gz
Visualisation	R 3.1.0 with packages: plyr rworldmap TeachingDemos ReadImages

Table 4.1: Software and Libraries

4.2 Data Source Formats

4.2.1 Tweets

The Experimental Tweets were taken from a JSON collection which had been created using the Twitter Streaming API. An example Tweet is shown in listing 4.1.

Listing 4.1: Example Tweet Format

```
{"hits" : {
    "total" : 7791,
    "max_score" : 1.0,
    "hits" : [ {
      "_index" : "geo_tweets",
      "_type" : "ctweet",
      "_id" : "138324985695490837898741517 3",
      "_score" : 1.0,
      "fields" : {
        "user.verified" : [ false ],
        "place.name" : [ "East Kootenay" ],
        "user.id" : [ 1869087644 ],
        "entities.urls.expanded_url" : [ "http://flic.kr/p/e5jPjx" ],
        "geo.coordinates" : [ 49.636608, -115.955261 ],
        "created_at" : [ "Thu Oct 31 20:04:11 +0000 2013" ],
        "place.full_name" : [ "East Kootenay, British Columbia" ],
        "place.country" : [ "Canada" ],
        "place.place_type" : [ "city" ],
        "place.country_code" : [ "CA" ],
        "entities.hashtags.text" : [ "Odessa", "Ukraine", "ordinary", "create", "site", "li" ],
        "user.friends_count" : [ 198 ],
        "text" : [ "For people from #Odessa, #Ukraine : Adidas #ordinary fan, of any ordinary clothing brand could #create a #site #li http://t.co/X42T1ny6Re" ],
        "user.followers_count" : [ 198 ]
      }
    }
}}
```

4.2.2 SenticNet Data

SenticNet is one of the emotion mining resources used. It is available in RDF format as shown in listing 4.2. The main attributes needed from this resource are pleasantness, attention, sensitivity, aptitude and polarity for word concepts.

Listing 4.2: Example Snippet of a SenticNet Entry in RDF

```
<rdf:Description rdf:about="http://sentic.net/api/en/concept/console">
<rdf:type rdf:resource="http://sentic.net/api/concept"/>
<text xmlns="http://sentic.net/api">console</text>
<pleasantness xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.686</pleasantness>
<attention xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.621</attention>
<sensitivity xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">-0.241</sensitivity>
<aptitude xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.335</aptitude>
<polarity xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.467</polarity>
</rdf:Description>
```

Due the use of a workaround concept extraction technique described in 3.3.2.3, a high access performance for SenticNet was needed and RDF was not expected to deliver satisfactory performance. Instead, the database was converted from RDF to NTRIPLES formatting using Rapper.

Listing 4.3: Example Snippet of a SenticNet Entry in NTRIPLES

```
<http://sentic.net/api/en/concept/console> <http://sentic.net/apitext> "console" .
<http://sentic.net/api/en/concept/console> <http://sentic.net/apipleasantness> "0.686"^^<http://www.w3.org/2001/XMLSchema#float> .
<http://sentic.net/api/en/concept/console> <http://sentic.net/apiattention> "0.621"^^<http://www.w3.org/2001/XMLSchema#float> .
<http://sentic.net/api/en/concept/console> <http://sentic.net/apisensitivity> "-0.241"^^<http://www.w3.org/2001/XMLSchema#float> .
<http://sentic.net/api/en/concept/console> <http://sentic.net/apiaptitude> "0.335"^^<http://www.w3.org/2001/XMLSchema#float> .
<http://sentic.net/api/en/concept/console> <http://sentic.net/apipolarity> "0.467"^^<http://www.w3.org/2001/XMLSchema#float> .
```

4.2.3 SentiSense Data

The SentiSense database is stored in XML format and assigns WordNet Synsets and their part-of-speech tags to emotions. Therefore, WordNet index files have to be used for matching the WordNet Synset Identifiers (SID) with actual words. Listing 4.4 shows a sample entry of the SentiSense XML file while listing 4.5 shows three example entries of a WordNet index file. The WordNet index file maps words to WordNet SIDs, which are used to lookup values in SentiSense.

Listing 4.4: Example SentiSense Entry

```
<?xml version="1.0"?>
<SentiSenseCorpus>
<Concept synset="SID-00152712-A" pos="adjective" gloss="lacking
    cordiality; unfriendly; &quot;a standoffish manner&quot;"
   emotion="disgust"/>
```

Listing 4.5: Example WordNet Index Entries

```
daily r 2 0 2 1 00081910 00179023
daintily r 2 1 \ 2 1 00306144 00306254
damn r 1 1 ; 1 1 00025805
```

4.2.4 EmoSenticNet Data

EmoSenticNet, which is used for emotion labelling, is presented in CSV formatting. Listing 4.6 shows three sample entries in the format:
`<emotion>,<anger>,<disgust>,<joy>,<sadness>,<surprise>,<fear>`

Listing 4.6: Example EmoSenticNet Entries

```
shame,0,0,0,1,0,0
sensation,0,0,1,0,0,0
self-esteem,0,0,1,0,0,0
```

4.3 Code Structure

As stated in section 4.1, the emotion mining program is written in Python. The following section lists the different classes and their purposes.

4.3.1 Modules and Classes

The code structure is depicted in figures 4.1 and 4.2 and consists of the following components:

- **EMiner**. EMiner is the main application class which connects Data Access, Tweet Parsing, Natural Language Processing, Sentics and Statistics modules.
- **TweetParser**
  - **TweetParserJSON** parses tweets from JSON format. The experimental corpora were generated from Twitter Streaming API and stored in JSON format.
  - **TweetParserCSV** parses tweets from CSV format. The evaluation corpus was stored in CSV format.
- **Statistics**
  - **Stats** is used to calculate corpus statistics.
  - **Evaluator** is used to compare predicted emotions with expected emotion values and calculates evaluation output.
- **NLP**
  - **NLPHelper** offers natural language processing functionality such as removal of hashtags.
  - **twokenize** is an external class used from the ark-nlp-POSTagger which supports Twitter-specific tokenisation of text.
- **DAO**
  - **SenticDAO** is a data access object used to access the converted SenticNet NTRIPLES database using SPARQL queries.
  - **SentiSenseDAO** is a data access object used to access SentiSense by mapping SIDs from WordNet index files to SentiSense entries.
  - **EmoDAO** is a data access object used to access EmoSenticNet, which is stored in CSV format.
  - **SignalsDAO** is a data access object used to access CSV files containing emoticons, acronyms and swear words.
- **SenticsParser** connects the three emotion databases and applies specific semantic rules to assign emotions to Tweets.

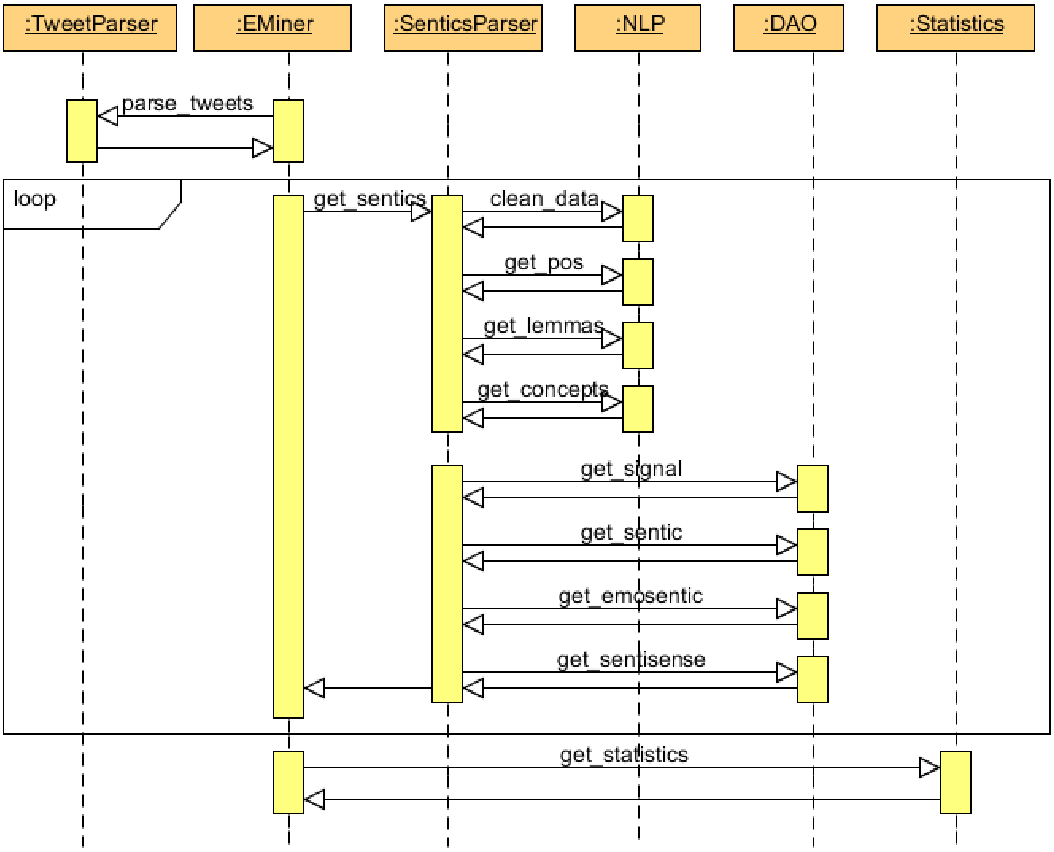


Figure 4.1: Sequence Diagram

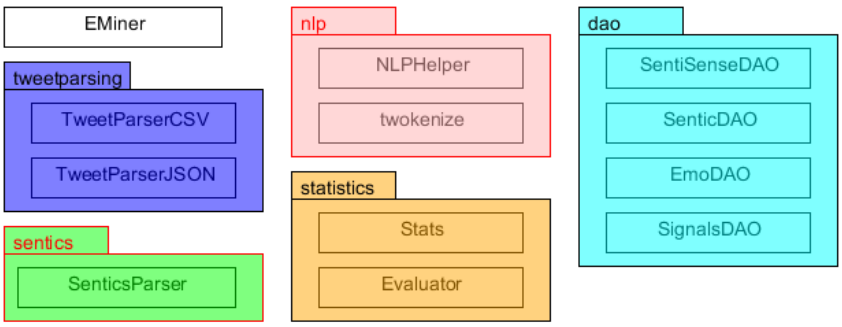


Figure 4.2: Package Diagram

CHAPTER 5

Experiment, Evaluation and Critical Reflection

In the following sections the proposed emotion mining approach is applied to experimental data considering possible use cases for supporting a horizon scanning activity. Furthermore, an evaluation is performed to test emotion mining accuracy and the applicability of this approach to horizon scanning. Finally, a critical reflection on the experimental results is done.

5.1 Aim of the Experiment

Based on the horizon scanning features described in section 2.1.1 the following scenarios were investigated as they were expected to show usable results:

(a) **Emotion Distribution Across Countries**. All Tweets of a specific topic and their emotion distribution across different locations.

(b) **Most Emotional Topic**. All Tweets of a specific location and their emotion distribution per topic.

(c) **Emotion Distribution Over Time**. Emotion distribution of all Tweets over time.

5.2 Data

Two Twitter corpora were used for this experiment as shown in tables 5.1 and 5.2.

5.3 Results

(a) **Emotion Distribution Across Countries**. All Tweets of a specific topic and their emotion distribution across different locations. The aim of this approach is to find differences in

Tweets:	41350
Dates:	1 Nov
Hashtags:	6012
Countries:	1 impact
URLs:	1494

Table 5.1: Corpus with UK Tweets from 1 Nov

Tweets:	7291
Dates:	Nov-March
Hashtags:	3156
Countries:	25
URLs:	2086

Table 5.2: Corpus with Ukraine Hashtags

emotion distributions for specific countries. A horizon scanner would scan the dataset for a country where emotion assignments for specific emotions differ from the other countries and generate a list of topics from the Tweets associated with these emotions. Corpus 5.2 was used for this experiment. Figure 5.1 shows that the distributions of emotions for the UK, Ukraine, Ireland, Canada and the US are similar. This occurs despite our initial expectation that Tweets from Ukraine would show a different emotion distribution. A possible explanation is that Tweets that are written in English are written from authors associated with the English speaking press and therefore convey similar opinions on Twitter.

	Emotion					
Country	Anger	Disgust	Fear	Joy	Sadness	Surprise
Canada	21,24%	12,95%	17,10%	32,64%	15,54%	0,52%
Ukraine	16,88%	12,24%	22,57%	37,97%	8,44%	1,90%
United Kingdom	20,86%	11,98%	19,82%	30,03%	13,91%	3,40%
United States	21,09%	9,38%	20,25%	35,45%	11,06%	2,78%

Figure 5.1: HeatMap of Emotion Distribution Across Countries

(b) **Most Emotional Topic**. All Tweets of a specific location and their emotion distribution per topic The corpus used for this experiment is described in table 5.1. All Tweets were analysed for the 16 hashtags that were mentioned the most and emotion distributions for these hashtags were examined. Similar hashtags, such as *#halloween* and *#HappyHalloween* were grouped together as they were expected to refer to the same topic. The results are illustrated by a heat map in figure 5.2 showing different emotional distributions for different topics. This allows for quick identification of the most emotional among the most popular Twitter topics. A horizon scanner may add the most emotional topic to his or her list of topics which is subject to further manual examination.

Emotion	BB7	bbcqt	excited	ff & FF	halloween, HappyHalloween, TrickOrTreat	jobs	Movember	nowplaying	ReleaseTheHounds	RespectGauaharKushal, Unfair Salman, wewantkushalback	trndnl	UK
anger		2,70%	15,63%		3,67%			1,82%	16,67%			3,03%
disgust		5,41%		0,82%	1,83%	3,13%	2,17%	1,82%	4,17%			3,03%
fear		2,70%	6,25%	2,46%	3,21%	3,13%	2,17%	2,73%	16,67%			3,03%
joy		27,03%	28,13%	23,77%	12,39%		36,96%	15,45%	20,83%			
neutral	100,00%	54,05%	46,88%	70,49%	70,18%	93,75%	54,35%	75,45%	33,33%	100,00%	100,00%	90,91%
sadness		2,70%		1,64%	4,13%			1,82%	8,33%			
surprise		5,41%	3,13%	0,82%	4,59%		4,35%	0,91%				

Figure 5.2: Heat Map of Emotion Distribution Across Topics

In this example from 1 Nov 2013, the identified topics could be considered to indicate the most talked about topics of the day. Despite the fact that we did not evaluate this assumption, the occurrence of topics such as *Halloween* and *Movember* in a top 16 list for 1 Nov 2013 is not expected to be random considering the dates of the extracted Tweets. The most emotional topic was *ReleaseTheHounds*, which is a TV show and therefore may be of a limited interest for a horizon scanning researcher. A closer look shows that *bbcqt* has a similar emotion distribution. `Bbcqt` is used when Tweeting about live political debates on BBC. Example Tweets that may be interesting are:

FEAR `Have you not heard. You can now watch court procedures. Why put it in the papers. #bbcqt`

SURPRISE `#bbcqt WHY do we need to switch?? The price of energy should be FAIR to start with no matter who supplies it!!`

JOY `Hurrah - such sense #renationalise #energy #nhs #trains #bbcqt`

JOY `#bbcqt #windturbines #solarfarms are the way to go! Clean #energy whats the alternative? A hideous factory? #NIMBY attitudes need to change!`

It may also be helpful manipulate the data by removing all neutral and positive emotions to get a complete picture of relative quantities of different topics and their negative emotional mentions. Figure 5.3 shows an example illustration in a tree map.

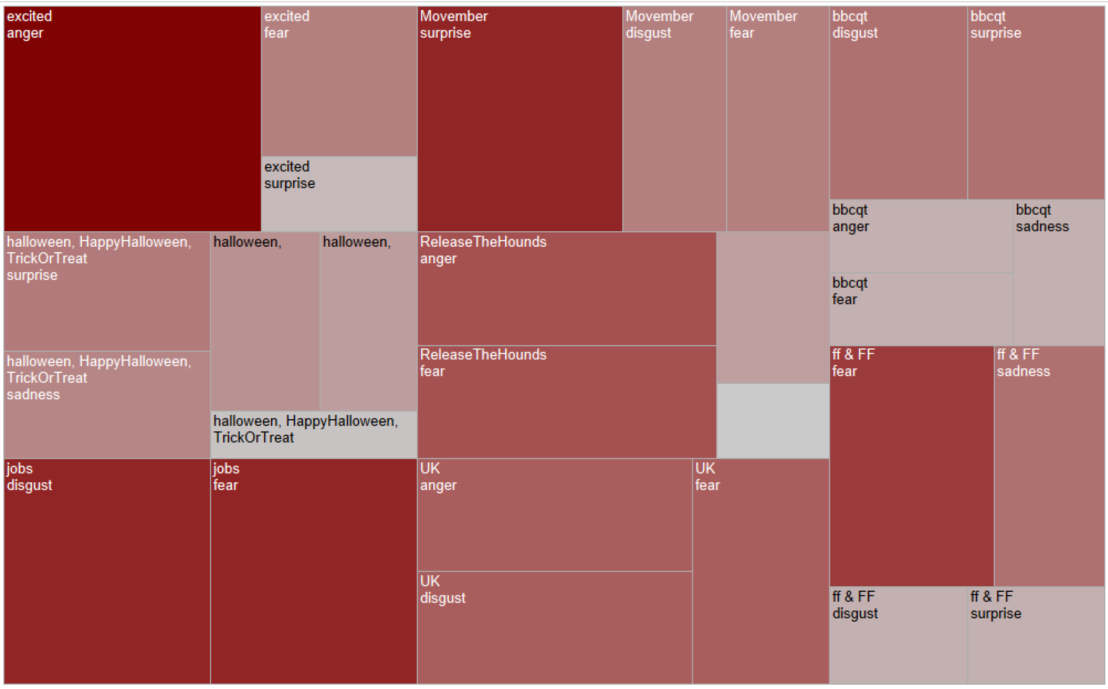


Figure 5.3: TreeMap of Top Topics and Emotions

A closer look at the absolute values for Tweet numbers shows that a corpus consisting of roughly 41350 Tweets may be too small for this type of analysis. As illustrated in figure 5.4, 15 most popular hashtags accounted for only 828 of the Tweets. This suggests that a large variety of hashtags or no hashtags at all are used in Twitter micro-blogging posts. Furthermore, 75% of the Tweets were classified as neutral as shown in table 5.3. Therefore, the total number of Tweets should be significantly increased to allow for authentic statistical analysis.

BB7, RespectGauaharKushal, Unfair Salman and 1 more	bbcqt	excited	FF & ff	HappyHalloween, TrickOrTreat, halloween	jobs	Movember	nowplaying	ORiginators	ReleaseTheHounds	trndnl	UK	Grand Total
111	37	32	122	218	32	46	110	26	24	37	33	828

Figure 5.4: Absolute Numbers of Tweets for Specific Topics

Emotion	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise
Percentage of Tweets	2.54	1.57	2.78	14.61	74.52	1.93	2.05

Table 5.3: Emotion Distribution of UK Test Corpus

(c) **Emotion Distribution Over Time**. Emotion distribution of all Tweets over time

Figure 5.5 shows how the emotion distribution for Tweets tagged with one of the hashtags shown in listing 5.1 changed over time. Looking for distinctive patterns, a decrease of Tweets classified with the emotion *Anger* can be detected, whereas the amount of *Fear* Tweets increases. This may be interpreted as the transition from a hot conflict to a cold conflict. More precisely, during a hot conflict there is more use of emotional language to express strong emotions, while a cold conflict is characterised by language indicating passive emotions such as apprehension and fear. A horizon scanner may interpret the increase of *Anger* Tweets as weak signal and specifically look at the corresponding Tweets. This could be followed by an extraction of topics related to these Tweets resulting in a list of topics which would be further interpreted by humans.

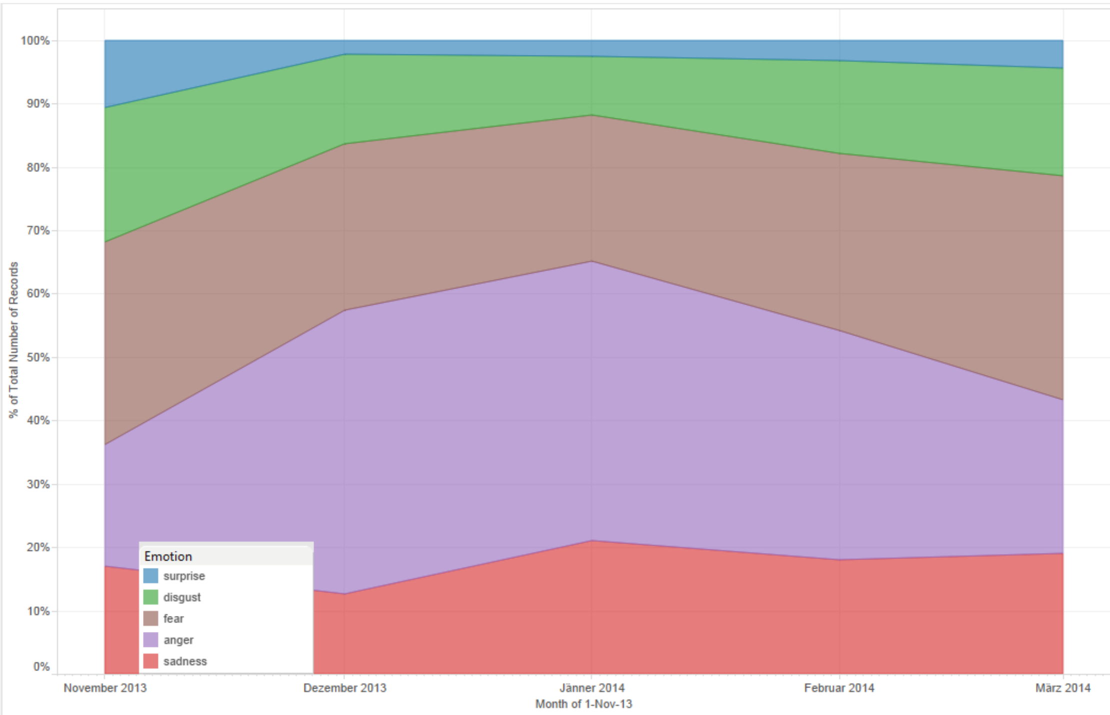


Figure 5.5: Changes of Emotion Distribution Over Time

Listing 5.1: Hashtags used for the analysis

```
\#ucraine, \#ucrainia, \#ukraine, \#maidan, \#euromaidan, \#
    kiev, \#donetsk, \#crimea, \#putin, \#prorussians, \#
    saveukraine, \#stopputin
```

5.4 Evaluation

Based on the state of the art analysis in section 2 and the design considerations in section 3.1, two main evaluation criteria for evaluating the proposed approach have been identified:

(1) **Emotion Mining Accuracy**. The precision and recall values should be comparable to other emotion mining approaches while being based on freely available resources.

(2) **Applicability to Horizon Scanning**. The results of the emotion mining approach should be applicable in a way which may support a horizon scanning process. Section 2.1.1 identified a criteria catalogue which will be subject to a qualitative assessment in this evaluation:

  a) Topics which newly emerge, leading to an increasing number of people talking about them.
  b) Topics relevant to society, which are not necessarily on policy agendas.
  c) Topics which occur in a specific domain and spread to other domains.
  d) Topics related to weak signals for threats, opportunities and trends.

5.4.1 Evaluation of Emotion Mining Accuracy

5.4.1.1 Evaluation Data

The twitter corpus provided for SemEval-2013 Task 2 [26] was used for evaluation purposes. It consists of 2320 Tweets, labelled as positive or negative. A random subset consisting of 600 Tweets was extracted and presented to two human annotators with different attributes as described in table 5.4.1.1. They were asked to assign exactly one of six emotions to each Tweet. If both annotators' Tweet classification is equivalent, the annotators agree and the classification is used as gold standard. For example, the Tweet *@appleTweets apple is about to lose my business. I recently bought this iPad may 31st... You come out with new iPad and discontinue ipad3!!!* was classified as `Anger` by annotator A and `Sadness` by annotator B and is therefore not included in the gold standard Tweets. The final set of gold standard Tweets was further altered so that there was an equal distribution of emotions.

ID	Gender	Age	University Degree
1	Female	26	Linguistics
2	Male	41	Psychology

Table 5.4: Human Annotator Profiles

5.4.1.2 Evaluation Process

The evaluation process consists of the following steps:

1. Download list of SemEval-2013 Task2 Tweets.
2. Extract 600 Tweets and store them in CSV formatting.
3. Present Ekman's six discrete emotions two both annotators A.1.
4. Give the same list of Tweets to both annotators. They receive a laptop with an open CSV file were each row contains a Tweet and numeric columns for each Ekman emotion.
5. Annotator 1 labels Tweets by choosing exactly one emotion for each Tweet by assigning the value 1 to it, all other emotions remain 0.
6. Annotator 2 labels Tweets by choosing exactly one emotion for each Tweet by assigning the value 1 to it, all other emotions remain 0.
7. Extract Tweets where both annotators agree (gold standard Tweets) and store them in a CSV file.
8. Check if gold standard Tweets have an equal emotion distribution.
9. Ensure equal emotion distribution.
10. Execute the emotion mining approach for these gold standard Tweets.
11. Compare the results to the gold standard by calculating precision, recall and accuracy using a python library *scikit-learn*.

During Step 8, it was found that the distribution of emotions was not equal. The most Tweets had been classified as *neutral*, a large number with *anger*, *joy* or *sadness* and a small number as *disgust*, *surprise* or *fear*. To ensure an equal distribution of emotions, additional Tweets were added to the evaluation corpus as follows:

- 100 Tweets from the Ukcraine test corpus described in 5.2
- 30 Tweets from *#disgust* on Twitter
- 30 Tweets from *#surprise* on Twitter

Steps 5–7 from the evaluation process were repeated for these corpora and in step 9 the final gold standard corpus was created. It consists of exactly 35 Tweets for each of the seven emotion classification values (Ekman's six discrete emotions + neutral).

5.4.1.3 Evaluation Results

The statistics module described in section 4.3 was used to calculate an evaluation output. Each emotion was assessed for its precision and recall values. Table 5.4.1.3 shows the evaluation results.

$Emotion$	$Tweets$	$Precision$	$Recall$
$NEUTRAL$	35	**0.23**	**0.49**
$FEAR$	35	**0.62**	**0.28**
$SURPRISE$	35	**0.67**	**0.10**
$SADNESS$	35	**0.41**	**0.51**
JOY	35	**0.48**	**0.60**
$DISGUST$	35	**0.60**	**0.34**
$ANGER$	35	**0.55**	**0.49**

Table 5.5: Precision/Recall Evaluation Results

Precision for an emotion x is the ratio of all Tweets correctly classified as x (True Positives) to all Tweets classified as x:

$$\frac{TruePositives}{TruePositives + FalsePositives}$$

Recall for an emotion x is the ratio of all Tweets correctly classified as x (True Positives) to all Tweets correctly classified as x + all Tweets incorrectly not classified as x (True Positives + False Negatives):

$$\frac{TruePositives}{TruePositives + FalseNegatives}$$

Concerning the 6 Ekman emotions, the result can be divided into two groups of emotions:

- *Joy*, *sadness* and *anger*, which show precision and recall values close to the 50% mark. For instance, *joy* has precision 0.48 and recall 0.60 which can be interpreted by stating that 48% of Tweets classified as *joy* were actually *joy* and 60% of all *joy* Tweets were classified as Joy.

- *Fear*, *disgust* and *surprise*, which show high precision (>60%) but low recall (<35%).

Concerning the *neutral* emotion, a precision of 0.23 shows that only 23% of all Tweets classified as *neutral* were actually *neutral*. Due to the fact that Tweets are only classified as *neutral* if no other emotion can be found, a dictionary based approach seems not be able to capture all emotions conveyed by written language.
In conclusion, the proposed approach meets the aim of being being reasonably accurate while using freely available resources. A critical reflection of the obtained results is done in section 5.5.

5.4.2 Evaluation of Applicability to Horizon Scanning

Section 2.1 identified different topic categories which can be indicators for emerging issues. This section assesses which of these categories of topics can be automatically detected by using the proposed emotion mining approach. The assessment is based on the experimental results described in section 5.3. If the proposed emotion mining approach covers at least one of the identified topic categories, it is considered as applicable to horizon scanning. This precondition allows it to be used in combination with other manual or automated tools to facilitate the whole horizon scanning process. The proposed approach covers two of the identified topic categories and is therefore considered as applicable to horizon scanning. See table 5.4.2 for details.

Topic Category	Assessment	Covered?
Topics which newly emerge, leading to an increasing number of people talking about it.	As the emotion mining approach does not assess novelty of topics, this indicator is not considered by the proposed method.	✗
Topics which occur in a specific domain and spread to other domains.	As this work does not propose an approach which looks at topics from a domain-sensitive perspective, this indicator is not considered by the proposed method.	✗
Topics relevant to society, which are not necessarily on policy agendas.	Based on the assumption of Klerx [23] it can be argued that an indicator for topic relevance is the occurrence of emotions in the written context of a specific topic. The experiment showed that the proposed approach generates an overview of the most popular topics and emotion distributions across these topics. Displaying this data on a heat map allows for quick identification of the most emotional topics which may be added to a list for further manual examination for horizon scanning.	✓
Topics related to weak signals for threats, opportunities and trends.	Weak signals may be interpreted as specific patterns of emotions occurring in micro-blogging posts. As the proposed approach allows for examination of emotions over time, patterns and corresponding changes can easily be observed. In order to show this, a Twitter corpus from November 2013 to March 2014 containing Tweets related to Ukraine was analysed. The experiment showed an increase of the emotion *anger* in time which turned into a decrease with an increase of *fear*. A horizon scanner would notice this pattern and interpret it as an indicator for a weak signal that may be an indicator for an emerging issue. Further analysis of the Tweets leading to this pattern would be a next step in a horizon scanning process.	✓

Table 5.6: Evaluation of Applicability to Horizon Scanning

5.5 Critical Reflection

The main criteria used for evaluation were emotion mining accuracy and horizon scanning applicability. This section critically analyses both aspects, focusing on implications for applying the proposed approach.

Emotion Mining Accuracy Precision and recall values were calculated in order to assess emotion mining accuracy. Although these are common measures for information retrieval systems, they have to be applied with care in the context of emotion mining. This is because there is no single truth for the main emotion conveyed by a Tweet. For comparing computed results with expected results, a set of gold standard Tweets was created by human annotators. This was done by manually labelling a Tweet with exactly one emotion. The following factors may influence the emotion labelling process:

- Psychological Definitions

  Psychologists have long differentiated between different feelings and ways of categorising them. These include mood, empathy, prototype and emotional episode [39]. The proposed emotion mining approach attempts to measure emotional episodes which are defined as events which fit pre-defined emotion labels. Human annotators may have different understandings of the meaning of the term *emotion*.

- Emotion Assignment Strategy

  There are different ways for manually assigning emotions to text, such as:

  - Guessing what emotion the author of a text really felt by assessing the meaning of a piece of text.
  - Checking for emotions conveyed by specific words and word combinations, e.g. *awesome* is likely to convey the emotion *joy*.

  The propose assign emotions to Tweets based on word combinations. Therefore, human annotators who act likewise in their emotion labelling process will generate a better result.

- Personal Experiences and Cultural Values

  Personal experiences may influence an annotator's view on emotions conveyed by specific statements. For instance, the Tweet *Yesterday I was drunk* may be regarded as positive or negative. The main emotion dictionary used in this work, SenticNet, would label this Tweet with *disgust*.

Due to these factors, two annotators with different ages and psychology/linguistics background were asked to annotate the Tweets and only Tweets where both agreed were used as gold standards. However, there is still room for misconception. Furthermore, the proposed approach is based on dictionaries which were created by humans which were also influenced by the factors described above.

The approach needed to be reasonably accurate in order to ensure its usefulness. When trying to quantitatively compare the evaluation results with existing approaches, it could be found that an average precision of 50% lies in between the results of the approaches discussed in section 2.2. However, a clear quantitative comparison with results of other approaches is not possible because of the following factors:

- Approaches differ in used emotion dictionaries and models.

  For example, UPAR7 (section 2.2.4), which has an average precision of only 27%, uses numerical values from 1 to 100 for each of six emotions. In comparison, the approach proposed in this thesis assigns only one discrete emotion to a Twitter-post. For example, in UPAR7 an emotion assessment may result in the following attributes: *anger: 20, sadness: 30, joy: 0, disgust: 5, surpise: 5, fear: 45*. In contrast, this approach proposed in this thesis only identifies one main emotion - in this case: *fear*.

- A corpus is highly context-sensitive. Test corpora used in existing approaches differ in size, origin and type of authors.

  For instance, EmoHeart, as described in section 2.2.1, uses 160 sentences from online diary-like blog posts. Diary posts contain personal thoughts and have very specific characteristics [2]. Therefore, text written inside diaries may differ widely from Twitter posts which are intended to be read by a wide audience.

- Different annotators are used to build gold standard Tweets. Sometimes authors annotate gold standard Tweets themselves and sometimes they use different kinds of experts. Rules of how many of them have to agree to form a gold standard Tweet also vary across approach evaluations.

- Different approaches perform emotion mining on different parts of text.

  For instance, 75% of the approaches presented in section 2.2 evaluate single sentences, rather than whole micro-blogging posts. The proposed approach focuses on analysing full microblogging-posts, as hashtag assignments are usually done for a whole Tweet rather than a single sentence. That is, in Tweets such as *Injured protesters are being carried to hospital. I have a bad feeling about all this. #euromaidan*, the aim was to derive an overall emotion, which could subsequently be attributed to the hashtag #euromaidan.

- Different evaluation measures which cannot be converted to each other, e.g. correlation and precision/recall.

  For example, SentiStrength, which uses separate positive and negative scales to assign emotional value to text, calculates correlation values. Their result is a correlation coefficient of 0.541 for positive emotion values and 0.499 for negative emotion values.

- Some approaches do not offer a complete description of their evaluation.

  For instance, SenticNet authors also proposed applications of their dictionary. However, evaluations for an application context similar to the one of this thesis have not been published.

Beyond Precision It may be argued, that an average precision of 50% is not sufficiently accurate to guarantee for usefulness of the proposed approach. However, emotion mining is highly context sensitive. Thus, different types of texts, such as political texts, would need different emotion mining resources in order to generate higher precision values. The proposed approach uses general context-independent dictionaries. Thus, it can be applied to all Twitter microblogging-posts, regardless of what kind of topics they cover.

For applying the approach to horizon scanning, having a way of detecting consistent patterns of emotion labels is at least as important as actually detecting correct emotions. Examples for such patterns are observing changes of emotion labels on Tweets over time and evaluating overall occurrence of non-neutral Tweets for assessing the relevance of a topic as potential indicator for an emerging issue. Detection the correct emotion is a non-trivial task both for a human and for an automated tool. This is because there is no single truth for the main emotion conveyed by a Tweet. Furthermore, awareness of certain precision values for specific emotions is crucial when analysing emotion mining results.

Horizon Scanning Applicability For assessing horizon scanning applicability, four topic categories which may be indicators for emerging issues were identified. Following that, an exploratory approach was applied, testing which ways of looking at emotion mining results can actually help in horizon scanning. Throughout the experiment, it was found that two out of four topics categories can be detected with the proposed approach. Therefore, it may be used to facilitate the horizon scanning process. It is to be noted that horizon scanning is still mainly done manually and can be hugely improved by introducing automatic methods such as the proposed approach. It introduces the novel idea of using emotional Tweet content when looking for indicators for emerging issues. Furthermore, it presents emotion mining results in a horizon-scanning-friendly format featuring attributes such as location, time and topic. This offers a number of possibilities for plotting the emotion mining results in order to facilitate a horizon scanning process.

These results show that the proposed novel approach is an appropriate tool for emotion mining and horizon scanning based on Twitter data.

CHAPTER 6

Conclusion and Future Work

6.1 Summary

The aim of this work was to design an emotion mining approach which takes into account Twitter characteristics and presents results in a format which may be used to support the horizon scanning process. The design science approach was used to (1) analyse state of the art in emotion mining, (2) design an emotion mining approach, (3) implement a prototype and (4) evaluate the results.

State-of-the-Art. An Analysis of current literature in horizon scanning showed that there is hardly any scientific work on automated horizon scanning. However, indicators for emerging issues, which are relevant to horizon scanning, could be identified. Furthermore, current literature gave clues about how emotion mining can be used for horizon scanning. The idea of using emotions as relevance indicator for topics which are being discussed on Twitter serves as foundation for this thesis. Current emotion mining approaches use a variety of techniques and resources. Examined approaches can be divided into keyword spotting and concept based techniques. While keyword spotting techniques rely on simply assigning emotional values to single keywords, concept based approaches go beyond single words and use word combinations, also called word concepts. Existing approaches further differ in their kind and extensiveness of natural language processing techniques as well as the weighting strategies used to assign emotions to words, phrases and sentences. There is a large variety of existing resources which can be differentiated by (1) the number of words or word concepts covered , (2) used emotion models, (3) the existence of a valence dimension indicating emotional strength and (4) public availability. Some existing dictionaries have been widely adopted, while others have been specifically designed for one single purpose and are not accessible. While the use of resources varies across different emotion mining approaches, their general structure was found to be similar.

Design and Implementation. The proposed design uses identified steps necessary to assign emotions to text:

1. Tweet Parsing, to parse Tweet data such as text, location and time.
2. Natural Language Processing, to translate Tweet text to a machine processable format, e.g. by replacing *don't* with *do not*.
3. Symbolic Cue Analysis, to look for swear words, social acronyms, interjections and emoticons
4. Emotion Mining, to use emotion dictionaries for mapping words and word combinations to emotions
5. Visualising, to illustrate emotion mining results together with location, time and topics.

The proposed design was implemented in Python, using JSON, CSV, XML and NTRIPLES as data formats for Tweet and emotion dictionary data.

Experiment and Evaluation. An evaluation was performed based on two main criteria: *emotion mining accuracy* and *applicability to horizon scanning*. Three Twitter corpora were used to evaluate the proposed approach: (1) a corpus containing 600 Tweets from the SemEval 2007 Conference, (2) a corpus containing Tweets originated from the UK on 1 November 2013 and (3) a corpus containing Tweets with one or more Ukraine related hashtags, such as *#euromaidan* and *#stopputin*. Corpus (1) was adapted to feature an equal distribution of emotions among Tweets and used for evaluating emotion mining accuracy. Results show an average precision of 50%. Corpora (2) and (3) were used to evaluate applicability to horizon scanning. An experiment was done where different ways of viewing emotion mining results were examined. This included showing emotion distributions over time and by location as well as listing the most popular topics along with their assigned emotions. It was found that a heat map of popular topics with their emotion distribution is most useful for supporting horizon scanning activities. However, although corpus (2) contained 41350 Tweets, the topic mentioned the most was only mentioned 216 times. This is the case despite consolidation of three hashtags (*#HappyHalloween*, *#halloween*, *#TrickOrTreat*) for this topic. As 60% of Tweets were classified as neutral on average, a higher number of Tweets would be necessary to gain usable results on emotion distribution for popular topics. The analysis of corpus (3) over time showed that changes of emotion distribution over time can be indicators of weak signals and are also relevant to horizon scanners.

6.2 Limitations and Future Work

Based on the analysis of other emotion mining approaches in chapter 2, several features are likely to improve the accuracy of the emotion mining approach:

6.2.1 Spellcheck

In order to detect misspelt words and word concepts, a spellcheck algorithm could be employed which corrects spelling mistakes. For example, *I kno ya* would be altered to *I know you*.

6.2.2 Negation and Irony

Simple rules could be used to handle simple cases of negation. An example is checking for the words *no* and *not* in bigrams. A dependency parse tree could further be used to detect negations that are not direct successors of their target words and relate them to the right target words. For example, in *I am not entirely sure*, the target word *sure* is described by the negator *not*. Dependency parsing could also be used to check for irony. However, this problem is still unsolved and difficult to handle as syntax offers only very few indicators for irony and sarcasm.

6.2.3 Boosting Emotions

Certain features of written text could be used to boost the emotional strength of a word. For example, hashtags could be seen as an indicator that the word which is followed by a hashtag is more important than other words. Another indicator for boosting emotions are modifiers:

- Modifier adverbs: only good, awfully optimistic
- Modifier adjectives: inferior quality

The characteristics of the modifiers could be used for boosting specific emotions.

6.2.4 Affective Grid Evaluation

An alternative way of evaluating an emotion mining approach is to use an Affective Grid [40]. The Affective Grid is a two-dimensional matrix, which allows emotional classification by degrees of arousal and pleasantness (see figure 6.1). A respondent would therefore not need to focus on discrete emotions when annotating Tweets. Instead, he or she would set a mark in a given matrix and define emotionality along two dimensions by doing so. These dimensions could further be mapped to discrete emotions and compared with computed values. However, this is not a trivial task and therefore Affective Grid evaluation is more likely to be accurate if compared with the exact two dimensions it uses itself. A way of evaluating the resulting matrices would be to assign numerical values to the mark and build correlations with computed emotion mining results.

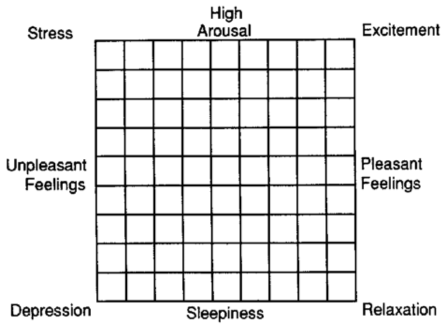


Figure 6.1: Affective Grid [40]

6.2.5 Horizon Scanning Applicability

The experiment showed that a heat map containing the most popular topics along with their emotion distribution is most usable for a horizon scanner. However, the experiment was only done with two corpora. In order to gain further insights, the following factors could be considered:

- **More data**. Less than 50.000 Tweets, which was not enough for statistical analysis as there is a high variation of topics among the Tweets. For instance, in our 41.340 Tweets corpus, the topic mentioned the most was only mentioned 216 times.
- **All hashtags**. It was found that a limitation of hashtags as it was done for the Ukraine-related corpus is not necessarily beneficial for a horizon scanner, as his primary aim is to find new topics.
- **All locations**. All experiments were done with Tweet corpora with only geo-location-tagged Tweets. Similar analysis should be done with all Tweets to assess differences.
- **All dates**. When examining emotion distributions over time, it could be interesting to look at daily averages rather than monthly averages.
- **Dynamic visualisation**. A dynamic visualisation could be implemented for allowing horizon scanner to drill down on a data visualisation.

List of Figures

List of Tables

APPENDIX **A**

Appendix

A.1 Appendix A - Evaluation Material

The materials used for the evaluation of the emotion detection accuracy of the emotion mining approach are attached here:

- Ekman Emotions description presented to annotators
- List of tweets where annotators agree

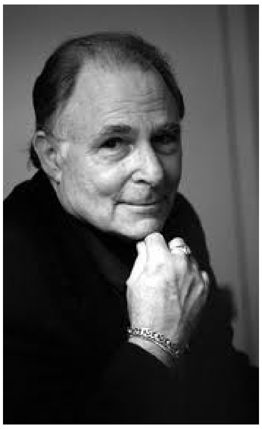
Ekman's Six Basic Emotions: List, Definitions & Quiz

In this lesson, we'll discuss Paul Ekman and how he discovered the six basic emotions universal across all cultures. You can view images and read descriptions of these emotions, and then you can test your knowledge with a quiz.

Taught by
Sarah Cobarrubias

We also recommend watching **Introduction to Emotions** and **Categorizing Emotions**

Who Is Paul Ekman?

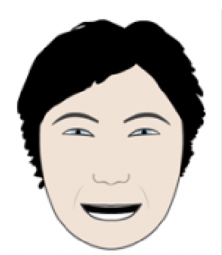

Psychologist Paul Ekman

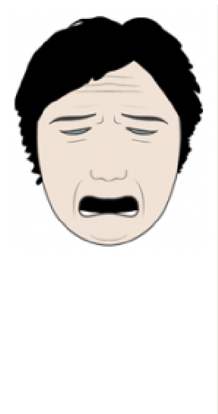
After his mother developed a mental illness and committed suicide, psychologist and behavioral scientist **Paul Ekman** dedicated his life to psychotherapy and helping people with mental disorders. He first began his research in nonverbal communication in the 1950s, developing systematic ways to measure body language. In the process, he discovered that, through empirical research, he could consistently identify facial expressions created by the movement of muscles in the face. And so, Ekman extended *his research to include facial expressions and their meanings*.

The Six Basic Emotions

Before Ekman hit the scene, it was widely believed (by anthropologists including Margaret Mead) that facial expressions and the emotions they represent were determined by culture - that people learned to make and read facial expressions from their societies. Ekman set out to test this notion in 1968. He travelled to Papua New Guinea to study the facial expressions of the secluded Fore tribesmen, where he learned that they could consistently identify emotion in facial expressions by looking at photos of people from other cultures, even though the tribe had not been exposed to any outside cultures.

It became evident, then, that *facial expressions are cross-cultural*; his research revealed that there is a *universal set of certain facial expressions used in both the Western and Eastern worlds*. This list of universal facial expressions, which Ekman published in 1972, comprises the **six basic emotions**. Take at look at the list, as well as images, definitions and muscular movements of these emotions, below:

Emotion	Picture	Definition	Facial Muscular Movements
Anger		Antagonism toward a person or object often felt after you feel you've been wronged or offended	Lowering eyebrows, tightening and narrowing lips, glaring eyes, tightening lower eyelids; less commonly, thrusting jaw forward
Happiness		Pleasant feeling of contentment and well-being	Smiling - pulling up corners of mouth, contracting large orbital muscles around eyes
Surprise		Feeling of upset or	Raising eyebrows high

		surprise at an unexpected occurrence	(which may cause wrinkles across forehead), opening eyes wide, dropping jaw so mouth is agape
Disgust		Intense displeasure or condemnation caused by something offensive or repulsive	Narrowing eyebrows, curling upper lip, wrinkling nose
Sadness		Feeling of unhappiness or sorrow	Drooping eyelids, lowering corners of mouth, pouting lips, downcast eyes
Fear		Feeling of apprehension caused by perception of danger, threat or infliction of pain	Raising eyebrows/drawing eyebrows together, tensing lower eyelids, stretching lips horizontally, mouth slightly open

Adding To The List

Continue reading...

Related Videos

Introduction to Emotions

Categorizing Emotions

Theories of Emotion

See more »

Psychology Courses Supplemental Lessons Popular Articles

Anthropology 101: General Anthropology

Psychology 103: Human Growth and Development

College & Career Guidance Courses

College Placement Exams

Entrance Exams

General Test Prep

K-8 Courses

Skills Courses

Teacher Certification Exams

Corpus	Tweet
SemEval 2007 Task 2	--i'm already excited for Winter Jam, and it's all the way in February!
SemEval 2007 Task 2	;) Seven Penny Stocks to Watch with Heavy Volume for Friday http://t.co/1TvlFXta medical marijuana inc
SemEval 2007 Task 2	.. & i probably won't even be on when Justin does his RT/Follow sprees cuz of the time difference. Plus, he may not even tweet as much (C
SemEval 2007 Task 2	...woah, Deezer. Babel only came out on Monday, can you leave it up for longer than a day to give slow people like me a chance?
SemEval 2007 Task 2	(...) Five of us conquering the world together by using our songs to bring happiness to people" - Kim Hyung Jun #2700dayswithSS501
SemEval 2007 Task 2	((I'm grounded from the internet until Monday, so talk to me while I'm still at school! -.-))
SemEval 2007 Task 2	[Tempo] Vanuatu quake | A strong 6.2-magnitude earthquake rattled the South Pacific island of Vanuatu Sunday, seismo... http://t.co/xoSnYLyj
SemEval 2007 Task 2	@__Aniko you think mr.Calle let practice with you guys ,cause I don't have practice tomorrow ??..
SemEval 2007 Task 2	@__AsToldByDeee OMG! lucky den! i was 12 when i saw Drake for da 1st time and it was my 1st Drake concert lmao i was 13 for Club paradise :)
SemEval 2007 Task 2	@__Takiyah hmm if this Friday is the 28 then it's away at Seaholm. Next home is hc :(
SemEval 2007 Task 2	@_BaileyRobinson She sat there this morning when Levi picked up her shirt and then lifted her paw to show she wanted to wear it today lol
SemEval 2007 Task 2	@_L33GOV_ I have the big Ipad3 I love it but for reading its too big my mini comes on friday
SemEval 2007 Task 2	@1DsHotTamale ~.~t u cant just tweet sum1 GUESS WHO IMMA SEE THURSDAY and NIT TWEET THEM BACK!! >.< lol
SemEval 2007 Task 2	@ABTBallet In the Upper Room Sat night was awesome! and standout Simone Messmer - awesome! Wish it was filmed...
SemEval 2007 Task 2	@Ahisraell Lies! I've been failing the math TAKS since like 6th grade man. :o
SemEval 2007 Task 2	@AIR_artists @an_artnews sadly this is affecting us too.Relief turned down,rateable value up..awaiting response about 2nd appeal from LCC
SemEval 2007 Task 2	@AliTweel Guess im a bit father back for classics "Inn of the 6th Happiness", "Pennies from Heaven", "To Sir With Love" :)
SemEval 2007 Task 2	@alive_nolove I put 'Do not take out or upload a 2nd time (for example Youtube, Vimeo, Tudou, Youku, etc)' its the same thing.
SemEval 2007 Task 2	@Allie_cat10 it's not Friday! And all of Marshall's talking when u can't beat out freshman team! They couldn't even score
SemEval 2007 Task 2	@ameliabrfeg ow god haha, going home at break tomorrow btw cuz my anutys down from York so going for a meal xx
SemEval 2007 Task 2	@amie_07 I ordered calendar from Book Depository and it arrived safely. So, this is my 2nd time and I can say Book Depo is looking good!
SemEval 2007 Task 2	@appletweets apple is about to lose my business. I recently bought this iPad may 31st... You come out with new iPad and discontinue ipad3!!!
SemEval 2007 Task 2	@Ashweee1D oh nice!! You should come to our show at The Lair on Dec 1st with Heroes For Hire, Masketta Fall, Forever Ends Here and WSTP! :)
SemEval 2007 Task 2	@asto_90 @danhill1987 haha oh yeah you're right, who have Arsenal got? Catch you tomorrow boys! Up the Albion!
SemEval 2007 Task 2	@avalard Have a good trip. See you tomorrow at the Jurys Inn?
SemEval 2007 Task 2	@Ayooitsdock your going to be the reason why I fail this HSPA tomorrow, I know it
SemEval 2007 Task 2	@BarackObama, Clinton, Panetta, Petraeus we will not #StandDown on Nov 6 or Nov 7 or Nov 8th. Do the right thing now. #WeWillNotLetThisGo
SemEval 2007 Task 2	@bbcburnsy: Loads from SB; talks with Chester continue; no deals 4 out of contract players 'til Jan; Dev t Roth ,Coops to Chest'ld #hcafc
SemEval 2007 Task 2	@bdd9 Every day from work, Monday had the Paley screening, Tuesday had to run errands, tonight auditions ran late. No break yet. :(
SemEval 2007 Task 2	@BenJBaur done and done! I haven't seen the last few episodes. I don't have Internet until Monday so I can't download them yet
SemEval 2007 Task 2	@BrownhillsBob double lock the doors tonight. Mrs Bob's unconscious self may go wandering, Nekkid & angry. Not sure Brownhills is ready.
SemEval 2007 Task 2	@calloway24 @DrakecoachDL @JCFSports Bias is prevalent at all levels. You don't want 2 find yourself on the Bubble on Selection Sunday #NCAA
SemEval 2007 Task 2	@CarlAStrauss I'm from Halifax, Nova Scotia. Not sure if we will go on March break, I'll have to see as it gets closer. Why?
SemEval 2007 Task 2	@CDixonLegs @Messengerprobz @dixon_messenger Colton is busy right now and may not see it. Did you send it to his mother and/or Traci Scarce?
SemEval 2007 Task 2	@Cheetoss_96 Ugly, are you taking the CAHSEE tomorrow?
SemEval 2007 Task 2	@CNBC didn't iPOOP just came out with 3rd gen IJUNK not to long ago? why the 4TH??? CUZ they CAN NOT compete with SAMSUNG, ASUS,NEXES.,,,..,
SemEval 2007 Task 2	@Connor_Howieson went to work on Saturday and someone had spray painted 'Kony 2012' on the side of royal high school haha poor effort
SemEval 2007 Task 2	@ConnorMcManusx it wasn't Tuesday it was the week before when Nat and chidgey were arguing I was like oh gad
SemEval 2007 Task 2	@couriermail I may not agree with Wilkie's ideas but he has proven he is a man of principle...It's a pity Peter Garrett cannot say the same.

SemEval 2007 Task 2	@crystallakemike Nobody knows if DRose will have the same game ever again when he does return. March seems awfully optimistic to me...
SemEval 2007 Task 2	@CW_network Bummed We don't get to see Arrow until Saturday because of the Chicago Bulls game.
SemEval 2007 Task 2	@DanielCharles96 ah. You may survive it. York may have a word with you about it. He's a dick. A fucking 5's too short for his liking
SemEval 2007 Task 2	@Darrellissa Sir, there are enemies in the White House.Something needs to be done TODAY not November 6th.Clinton,Obama,Panetta,Petraeus,Lamb
SemEval 2007 Task 2	@DJDannyF never listened to the 1st. hard to ignore the 2nd working at HMV! but like most things, just got overplayed. agree on the 3rd
SemEval 2007 Task 2	@Dont__KAY_me omg same I was reading it in school after PSSAS and I just sat there crying
SemEval 2007 Task 2	@DonzThaDon @BlueVino May not have. But I just know appeared to be better in early 80s before Reagan fired Air traffic controllers.
SemEval 2007 Task 2	@Doug_Zamensky @SuzziFiazco @ianasbury #HARD #dayofthedead weekend just got a bit weirder. Chateau Marmont on Sunday? What?!
SemEval 2007 Task 2	@DrPravinTogadia I condemned both the measures but relating it to religion is cheap. If done in Jan then Holi, if done now, then Diwal
SemEval 2007 Task 2	@fredmacpherson We've lost HQ to Wills+Kate+the Olympics till mid September. I'll have words. Party on Boris Bikes tho...
SemEval 2007 Task 2	@GodivaGirlBH I cancelled with Amazon as they weren't going to post until the 29th! Went with Book Depository instead, still waiting
SemEval 2007 Task 2	@Gonzalexx I'm sorry I lack the drawing skills to pay homage to such a great illustrator as Jean "Moebius" Giraud on Tuesday Sketch 2.
SemEval 2007 Task 2	@hamilton_elle swing by Hollywood Tans if its too hot for the Dubai sun.
SemEval 2007 Task 2	@Hyperace I don't know why they haven't put Siri on the iPod 4th Gen.
SemEval 2007 Task 2	@ingrid_wien Asa and Kim bought me a Swarovski rose for Mothering Sunday. I haven't a clue what to do with it. It is still in the box.
SemEval 2007 Task 2	@j9neverlose Traveling the whole circuit? Cleveland can be pretty dismal in March....unless you're drinking w me for my bday on St Pattys
SemEval 2007 Task 2	@JeremyOnMarz just be tryna hate on ma' 2nd babydaddy Kirko Bangz.
SemEval 2007 Task 2	@JJWinter62 dancing round mate. Just shocking. You will have to try and find it on Internet tomorrow. Honestly bad bad bad boxing
SemEval 2007 Task 2	@JoeandEvan Ev, u went from 3rd in the east to not making the playoffs, Deron will miss 2-4 days of practice lol
SemEval 2007 Task 2	@justindsweeney BYU vs GT will re-air Saturday at 9am MT. We did not get VOD rights so it won't be archived on http://t.co/NnQCa8Am.
SemEval 2007 Task 2	@Kdizzle3434 Dude, it was sad saying goodbye to the PH last night. I'll always remember my 1st St Pattys when we stayed up all night..
SemEval 2007 Task 2	@KeiJhoHyun28 oo nga eh :(may picture sana ako with PREPIX!!!!
SemEval 2007 Task 2	@kellehn He is hip hop Sidney Crosby to me. May be the best at what he does but SUCH a whiney self important bitch that I don't give a damn
SemEval 2007 Task 2	@kenzzzzxo so the thing next Thursday isn't free, you'd have to pay $15 to get in since you don't go to UMBC :/ and it ends at 11:30
SemEval 2007 Task 2	@kris1179 If Spurs hadn't got that result mate you guys would have been 1st up. Have you heard SAF moaning. 4mins injury time an insult
SemEval 2007 Task 2	@LadySmaug Can't hover over link to see file name? Tudou doesn't have playlists and this particular person only had the 1st episode anyways.
SemEval 2007 Task 2	@lewiscrs Didn't Leverkusen reject a 20m offer for him in the summer? I fail to see how his stock may have dropped since then
SemEval 2007 Task 2	@LiebeTreueBrai lol Idek what phase my shit is..I just know morris hall 3rd floor hell
SemEval 2007 Task 2	@LL_Cool_E: DMU 1st years are boring you know...we may need to teach em - Fuck 'em, fuck 'em I'm screaming FUCK 'EM!!!
SemEval 2007 Task 2	@luvaprilrenee I was speaking in general..but Romney and Obama full of bullshit..put Dennis Kucinich on the ballot then I'll be 1st in line
SemEval 2007 Task 2	@lzone Yeah I was told 7-10 days, but then @kakaty found article stating it could be by Friday in Shaker. This sucks.
SemEval 2007 Task 2	@mariam02 japan fukaoka tix will be released on 24 nov. Im doubtful that we can get the tix coz the VIPs will have priority.
SemEval 2007 Task 2	@MattMahony86 They had a bit on E News the 29th with Adele's announcment and my dad called her a slut and not a good role model. He was
SemEval 2007 Task 2	@Mc_Squ1zzy @Fresh_Prince01 -__- shid i may not be able 2 take screenshots of it but everybody @'s me. Peyton Siva kik'd me this morn #NoLie
SemEval 2007 Task 2	@MeekMill just went to HMV and meek mills album don't come out till Monday #pissed #dreamsandnightmares #dreamsandnightmares
SemEval 2007 Task 2	@meow_x Unfortunately, not yet. Wanted to go to Revenge at Paleyfest in March, but couldn't make it. :(
SemEval 2007 Task 2	@MIPrepZone just driving past Groves. Scoreboard says 22-15, Seaholm... May not be accurate-I'm not at the game.
SemEval 2007 Task 2	@MittRomney <U>Wake Up Fox, not going there>W Bush years, Dennis Kucinich speaks @ DNC on Aug. 26, '08 WAKE UP AMERICA http://t.co/s8b5cfN5
SemEval 2007 Task 2	@moanajkidd The massacre on March 11 in Karm al-Zaitoun was similar: 25 children killed with knives, 20 women: https://t.co/hxc5Kexv
SemEval 2007 Task 2	@Monzzz330 were fucking going to Taco Bell tomorrow

SemEval 2007 Task 2	@MorgaineYou When i saw the tsunami on March 11 on NHK, it put tears in my eyes!!!
SemEval 2007 Task 2	@Muccas Sylvia, I can't make Thursday evening I have an event to go to. Can do it Saturday if that's ok? #norestforthewicked
SemEval 2007 Task 2	@mukurowl we don't even have naruto tomorrow just khr, bleach, and op. the negatives outweigh the positives.
SemEval 2007 Task 2	@nickelodeontv May you please show How To Rock from episode 1 and so on.....i cant blieve theres no more how to rock
SemEval 2007 Task 2	@Nike_Wearer the ACC had a team ranked 3rd guy. Florida State. Until they lost to NC State. ACC so underrated.
SemEval 2007 Task 2	@NotDanica7 If you're grandmother is your sister, you may be a #Redneck. Sorry, I sounded like Jeff Foxworthy. Please don't block me.
SemEval 2007 Task 2	@NotTheGolfer Could be similar to DRose as the 8 seed his 2nd year in. Don't know if the Cavs can make a jump like the Bulls to 1, but still
SemEval 2007 Task 2	@OoTEREoO Hey what's the status of Kina Grannis and her concert? If the rain is still this bad we might not have a show tomorrow...
SemEval 2007 Task 2	@OrangeBob77 I haven't made game-by-game predictions, but I think it'll take a few breaks to get to 27-28 wins before Selection Sunday.
SemEval 2007 Task 2	@PaigeTaylor_93 I think St patrick would be disapointed if you didnt get a snake up there for the next paddys day..think its on a thurs too?
SemEval 2007 Task 2	@pjforest0712 What a shame about Punchestown. Unlikely to race at all tomorrow. I had 2 out as well in Quevega and Lietenant (2 miler!!!)
SemEval 2007 Task 2	@ProfOfGoonism September 22 is a Bastard Bearded Irishmen half way to St. Patty's Day party at Mulligan's. I'm going and Erika's going
SemEval 2007 Task 2	@REAMER_72 I think Sean Parker is a fuck. He may have a point about it being boring. But the guy is a drug addict idiot that cant hold a job
SemEval 2007 Task 2	@RevBhoyULTRA i'm more upset by the racist, stereotypical shit that every fucking store under the sun sells around St. Patrick's day.
SemEval 2007 Task 2	@Rock_Kendy but the 12th man is still the 'Lionel Messi' for the manc scums unfortunately
SemEval 2007 Task 2	@Ronnie2K I'm sick of playing the Pacers & Gerald Green playing like the 2nd coming of Michael Jordan smh
SemEval 2007 Task 2	@sandralaville Sounds like April may know the person that has taken her but just hope we DON'T have another Tia Sharp case on our hands!
SemEval 2007 Task 2	@sandyboxx @trafiuddin @sawsjabr @amreen_taher ya which plans lol? And btww amreen can't even come Friday to the MSA event she has a bday :(
SemEval 2007 Task 2	@SAVE_US_GSM So Taker/HBK had the 2nd to least amount of blood? That shocks me cause HBK was in pool of his own blood at the end.
SemEval 2007 Task 2	@SaxbyChambliss @foxandfriends @datakcy What did Petraeus tell the Intel Comm. Sept 13 and why did @SenFeinstein blame the video?
SemEval 2007 Task 2	@shazzer42000 Poor match ended their winning streak. They also lost to Leverkusen at home for the 1st time in 24 years.
SemEval 2007 Task 2	@shelboxoxo I'm going to Cuba buuuttt I doubt I will be doing what y'all will #13.1 Sunday
SemEval 2007 Task 2	@SheStayMIA lux was super dry even with Lebron & Saturday I went to Cafe Asia, super lame!!! I'm getting too old for this shit
SemEval 2007 Task 2	@ShotMyLover then it goes to show that all you do is run your mouth but you can't ever come to our faces come to allderdice tomorrow bitch
SemEval 2007 Task 2	@Skinny_Minix3 retard you don't take Hspa until march cause you're a jr
SemEval 2007 Task 2	@sqorgar and then there are shows that durdle around between neat and terrible/generic the 1st season then get cancelled (Terra Nova)
SemEval 2007 Task 2	@Swish_KK32 and if ur amywhere near kansas state and dont come watch us play in the preseason NIT on nov. 12th im gonna fight u for that too
SemEval 2007 Task 2	@T_MoniqueXOXO They not for sports wit yo hatin ass lol I'm smart af. Yu just focus on friday and them practice Ogt's
SemEval 2007 Task 2	@tabhaverkamp that's horrible news that UMBC will wait until 6 am tomorrow to let us know about closing. Hhhhmmmm
SemEval 2007 Task 2	@TEEN_TOP oh noooo... i am really confused of chunji's age.... is he 20 or 19???? i just tweeted and said "happy 19th birthday to chunji" xD
SemEval 2007 Task 2	@Tina__Tubs Does anyone else find it odd #McCann jumped on April Jones bandwagon,but didnt say word about Tia Sharp where FAMILY involved ?
SemEval 2007 Task 2	@tonyrusciano Hearing Taco Bell, makes me want Taco Bell. I hate you now. Can we all go to Taco Bell tomorrow?
SemEval 2007 Task 2	@tswift_NZ And I don't think it helps that tomorrow is Labour Day haha. Yup, fingers crossed that some stores put it out!
SemEval 2007 Task 2	@turbohat a few speculating that the 3 had heat/power issues & 4th gen resolves this. But I have heard nothing around these probs on iPad3
SemEval 2007 Task 2	@unpredictagwapz get the new frequency, the phone you have now may not work on their LTE network anymore. But HSPA is always the same
SemEval 2007 Task 2	@valderie There was too much Grammy, I should've known we'd be sad at the end. Ps are you going to Dashboard tomorrow?
SemEval 2007 Task 2	@venomdjsk vampires dont sparkle.... look at.. that Moebius guy from new york (i havnt watched spiderman for a while....) he hates the sun
SemEval 2007 Task 2	@verge:Galaxy S II can be wiped by just clicking a link, other TouchWiz devices may also be vulnerable http://t.co/rJT2wr0u @SamsungMobile
SemEval 2007 Task 2	@vtti right?! That man is a U.S. Senator who sounds likes the most annoying kid in 8th grade home room.
SemEval 2007 Task 2	@Waka_BacaFlame I'm already on my 3rd year at Colton it wouldn't make sense to graduate from gt after spending all this time here >.<

SemEval 2007 Task 2	@WHunterWest @JonathanTepe might wanna check on me if we lose tomorrow... Might pull something similar to that Kony guy in the streets
SemEval 2007 Task 2	@Zwelinzima1 could have fooled me. I thought the march was about trying to win back favour with workers... Cosatu is fast losing credibility
SemEval 2007 Task 2	@Zwelinzima1 do we really have to march every time we have issues to deal with? Marchers clash at Cosatu rally http://t.co/JN9zwWMp
SemEval 2007 Task 2	#Bills Mario Williams not sure if he'll practice today (wrist) expects to be more effective Sunday. @wgrz
SemEval 2007 Task 2	#CNET - Just bought my 1st iPad, iPad3, feeling real burned, mad, about iPad4 so soon. Grrr. REALLY mad! Don't even care about mini now,
SemEval 2007 Task 2	#fakta on the 8th of Mar 191 Cao Cao attempted an assassination on Dong Zhuo; an evil tyrant that use the last Emperor of Han as a puppet.
SemEval 2007 Task 2	#HeadiesAll of you famzing live at the #Headies and dnt knw where Eko Hotel n suite dey..May ur twitter connection be bad.#GodPunishU
SemEval 2007 Task 2	#noagenda Russian NGOs to lose American funding as USAID closes. (Euronews video): In July, the Russian par... http://t.co/8XZnp2At #itm
SemEval 2007 Task 2	#UVa wide receiver Tim Smith says he'll be able to play in Saturday's game against NC State.
SemEval 2007 Task 2	1st debate showed emperor has no clothes...Wonder how the court jester, shoeless Joe will do against Ryan? #gop
SemEval 2007 Task 2	1st tour I was on,my brother was working carrying the bags.he'd stand on the side of the stage & shout out "I HATE YOU!" lol. miss u Miyagi!
SemEval 2007 Task 2	2 tires looked to be a bad decision for Knaus. Went from dominant to struggling to stay in top 10 in 3 laps. Now 7th just in front of 5, 29
SemEval 2007 Task 2	2nd win today, 2/1.. Lets keep it up guys ! Next game on saturday at Helsinki..
SemEval 2007 Task 2	3rd season Dance Academy means to me : going crazy! I love it and can't wait :)))
SemEval 2007 Task 2	7th lie: Sameera said that Nabeel's marches threatened police to use Molotov cocktails & iron rods ! she's not ashamed of pure lies #Bahrain
SemEval 2007 Task 2	A Cape Town woman died on the eve of her 21st birthday when she was knocked down on the N2 highway, the Cape Argus reports today
SemEval 2007 Task 2	A Florida player would never get away from a cop with pot and a gun in his car. Michael Dyer did - then got dismissed - http://t.co/yWYAOnVr
SemEval 2007 Task 2	A year ago tomorrow changed the ICONiac fan base forever :(
SemEval 2007 Task 2	Aaaaahhhhh dam! I am out of my favorite Tea! Damit I have to wait until tomorrow to buy it now :(
SemEval 2007 Task 2	About 400 bricks just cascaded from a building on 2nd between 42nd & 41st onto water system of Hilton Hotel / Tudor Hotel. Scary
SemEval 2007 Task 2	About to go get this lift, run and throw in. Jr. Pro Day on Monday then it's back to SA to get ready for the season!!! #HogNation
SemEval 2007 Task 2	According the BNP's website they are bussing people down from Sheffield for the "BBC Paedo" protest at Broadcast House tomorrow.
SemEval 2007 Task 2	Activists in Deir Ezzor captured this image of Musab Bin Umair Mosque after regime forces set it on fire Wednesday. http://t.co/MRcoprCE
SemEval 2007 Task 2	Ahhhh the next few weekends are going to be messy! First off - this saturday: Powder Room at the Gladstone Hotel... http://t.co/C04KQKNr
SemEval 2007 Task 2	Ahhhhh I can't wait til Laidback Luke drops show me love & speak up tomorrow! Last time he dropped them Sooo hardware!!! I love him!!!!
SemEval 2007 Task 2	AirAsia aborts deal to buy Indonesia's Batavia Air - Southeast Asia's top budget carrier AirAsia said Monday it has ... http://t.co/WLNFwyHS
SemEval 2007 Task 2	All the kids who talk shit about Northglenn caue they graduated are the same ones going to the game tomorrow.
SemEval 2007 Task 2	And you say Derrick Rose isn't good because of the ACL but Dwight Howard is the 3rd best player in the league coming off major back surgery?
SemEval 2007 Task 2	Andy Lee punts it down to the 14 after Crabtree couldn't come up with grab on 3rd down.
SemEval 2007 Task 2	ANDY MURRAY CRASHES OUT Andy Murray suffered a surprise early exit from the BNP Paribas Open in Indian Wells. http://t.co/V7c1Dmjt
SemEval 2007 Task 2	Anxious for the one next Wednesday I need to get to the bottom of this Bobbi Kristina and Nick thing
SemEval 2007 Task 2	Assad continues to bombard #Homs, 4 the 17th consecutive day, mobile ntwrks off 4 the 4th day, mjr battles around Baba Amr #Syria
SemEval 2007 Task 2	AUG 9TH St Practice Day. We haven't reached St Patrick's Day yet the snakes are still in Ireland - come in and get bitten! $4 Snake Bites
SemEval 2007 Task 2	Awkward when Marni in Halloweentown completely changes people between the 3rd and 4th movie #imsickandnothingison
SemEval 2007 Task 2	Barcelona face a defensive crisis when they host Granada on Saturday, with Carles Puyol (knee) and Gerard Pique (foot) both sidelined.
SemEval 2007 Task 2	Barcelona's Carles Puyol out 8 weeks with dislocated elbow suffered in Champions League win Tuesday
SemEval 2007 Task 2	Bayern Munich coach Jupp Heynckes insists his side won't be fazed by the shock loss to Leverkusen on Sunday - http://t.co/E0ctlsyA
SemEval 2007 Task 2	Bears suffer a 2-1 loss at SDSU tonight http://t.co/gVjbkrAk next up - No. 4/5 UCLA on Sunday in Los Angeles #GoBears
SemEval 2007 Task 2	Becoming dangerously obsessed with one of tomorrow's Melodifestivalen songs. I'm sure you can see why. http://t.co/UItTiPrg
SemEval 2007 Task 2	Beer, xbox, Undertaker: The Streak DVD. Fuck you Saturday night.

SemEval 2007 Task 2	Ben Howland implies he has no interest in #Huskers hoops coaching job. See LA Times ... http://t.co/nXJ5I4kx
SemEval 2007 Task 2	Billy Gillispie is awful. Nothing's worse than him being your coach. Unless it's losing to him in the 2nd round of the 2007 NCAA Tournament.
SemEval 2007 Task 2	BNP opposes Companies Act change: The BNP on Tuesday asked the government to drop the move to change Companies A... http://t.co/AVRh5J4g
SemEval 2007 Task 2	Bolton Wanderers have sacked manager Owen Coyle following a poor start to the Championship season, which sees them lying 18th in the table.
SemEval 2007 Task 2	Bolton Wanderers have sacked manager Owen Coyle following a poor start to the season. The club currently sit in 18th in the Championship.
SemEval 2007 Task 2	Bolton Wanderers, who are 18th in the Championship after winning only three of their 10 matches, have sacked their manager Owen Coyle.
SemEval 2007 Task 2	Book Depository may be cheap but the inferior printing quality of some of their products is almost not worth it.
SemEval 2007 Task 2	BREAKING NEWS: #Arrow will not be seen tonight in Illinois or Indiana due to the Chicago Bulls opening game. It will air on Saturday. :(
SemEval 2007 Task 2	Brilliant Burns stops Mitchell in 4th: Kevin Mitchell's bid to win the WBO lightweight title ended in failure on... http://t.co/GZxRbVmM
SemEval 2007 Task 2	Browns r cursed..."@evansilva: Chris Mortensen reported that the #Browns fear SLB Scott Fujita's career may be over due to a neck injury"
SemEval 2007 Task 2	Btw fuck Durant for going to the OKlahoma game Saturday!! You went to Texas!!! #LonghornForLife
SemEval 2007 Task 2	By the way Tia Sharp was buried today You may have failed to notice in all the fuss over a dumb posh tart getting her tits out.
SemEval 2007 Task 2	C'mon Sean Parker, step it up! Little ass nigha
SemEval 2007 Task 2	Calle stupid but I really don't feel like going tomorrow :c
SemEval 2007 Task 2	Can't believe how many hostels in Dublin are already booked up for St Patrick's Day!! 21st Birthday plans going down the pan =(
SemEval 2007 Task 2	Can't sleep. Got a lot on my mind. If I can't get to the WGC tourney in Russia, I really need to fight in Dec/Jan... Gotta stay busy!
SemEval 2007 Task 2	cant go to hot springs for the band marching contest because i didnt march at homecoming
SemEval 2007 Task 2	Cant sleep. Full mind. Watching "The Announcement" for the 5th time ft. @MagicJohnson #handsdown 1 of the most emotional sports stories ever
SemEval 2007 Task 2	Carles Puyol could be out for up to 6 weeks after straining knee ligaments. Further tests on the Barca man tomorrow - http://t.co/RfNj8osA
SemEval 2007 Task 2	Carles Puyol will miss the El Clasico against Real Madrid on Sunday after dislocating his left elbow yesterday. #LaLiga #LigaBBVA #FCB #RMFC
SemEval 2007 Task 2	Carnival Tomorrow! Wanted to rest but I don't think I could resist the temptation !
SemEval 2007 Task 2	CASHEE on Tuesday & Wednesday, math test on Thursday, an essay due Thursday too to top it all off rain all week! WTF!
SemEval 2007 Task 2	Casinos in Atlantic City closed since Sunday because of Sandy. What did the pathological gamblers do ?
SemEval 2007 Task 2	Cena and dolph will make 0 sense bc Ziggler will end up losing, like usual. Then he may cash in Who cares about tht if he loses 2 cena?
SemEval 2007 Task 2	Chandy rules out ban on toddy: Chief Minister Oommen Chandy said here on Wednesday that the government was not i... http://t.co/gTTA9rAq
SemEval 2007 Task 2	Charges filed in Calif. shooting that killed 3 - A convicted felon was charged Monday with the murders of three peop... http://t.co/VoLuhZe7
SemEval 2007 Task 2	Chuck Close is creating mosaics for the 86th st. station of the 2nd ave subway. This makes me hate the project a lot less. Well played, MTA.
SemEval 2007 Task 2	Come on iona just cancel class for tomorrow already
SemEval 2007 Task 2	Coming up in your Weekend Argus tomorrow: Kalk Bay waves cause some damage in the area.
SemEval 2007 Task 2	Congressman Dennis Kucinich (D-Ohio), who lost in the recent Ohio primary election on March 6th, due to... http://t.co/Z0q9fxP7
SemEval 2007 Task 2	Could someone please explain to me why my DMU account is still working? I've graduated!! I do not want to attend a seminar next Thursday.
SemEval 2007 Task 2	Dammn yung Tracy McGrady goin to play in China really wish injuries didnt slow him down was one the best in the @NBA my 2nd favorite player
SemEval 2007 Task 2	day 3: still got no power and takin like the 3rd cold as shower fml. fxckin freezing. reminds of peru lmao # http://t.co/FG3EyKp1
SemEval 2007 Task 2	Daylight savings time ends this weekend. And I don't know what this means for the timing on F1 on Sunday
SemEval 2007 Task 2	Daylight Savings Time is Sunday folks.. and we "fall" back.. maybe that will give some of you lethargic losers a reason to be early to work.
SemEval 2007 Task 2	Dear Gordon Merchant, to quote Peter Garrett, may I just say, "your Dreamworld is just about to fall...".. Karma biatches!! Suck my balls!!
SemEval 2007 Task 2	Dear Naked Cyclists, I will not support your campaign if you are putting your sweaty bits on Boris Bikes! http://t.co/TtnyoPj2
SemEval 2007 Task 2	Deir Ezzor :: Colonel Mohammed Al-Abdullah, from the 4th Division, and the Lt. Recruit Feras Mohammed Al-Abdullah have defected #Syria Ne...
SemEval 2007 Task 2	Deron just put me on punishment becuz I said I may be sleep by the time he may head to Rhons
SemEval 2007 Task 2	Derrick Rose may be the worst at doing press conferences, he can't talk to save his life

SemEval 2007 Task 2	derrick rose, you eat pussy and break ya neck modasucka loll "black friday" @michaelblackson http://t.co/90vpw3vW
SemEval 2007 Task 2	Do Endsmeat? Sharon Stone Rushed To The Hospital! - Oh no! Sharon Stone had a health scare on Saturday when she was ... http://t.co/EthDVGiF
SemEval 2007 Task 2	Don't be too upset about the lack of football this weekend, the soccer game against Seaholm is only a few days away! Tuesday. Home. 7.
SemEval 2007 Task 2	Don't wait til the last minute to book your booths and get tix for Soundboard on March 11th...you will live to regret it!
SemEval 2007 Task 2	Don't want to go to HSPA class tomorrow so I'm not going .
SemEval 2007 Task 2	Don't want to home tomorrow #omfg iHATEschool-.- Cuba is awesome...
SemEval 2007 Task 2	Download Fast Ipad Hd Video Player Guide: Do you have to shy away from looking to be whipped? You may not see at... http://t.co/ic3qKJ21
SemEval 2007 Task 2	DTN Cricket: Pattinson may be saved for A tour ahead of Ashes: James Pattinson may not make it to the ODI tour o... http://t.co/9nt69Ifk
SemEval 2007 Task 2	Eden Hazard's through ball to Juan Mata for the 3rd goal against Spurs - are you serious???!!! #onetouch #instinctive #class
SemEval 2007 Task 2	Emily Murray's @guardiannews "What Cherie Blair doesn't understand about yummy mummies" http://t.co/3NLedEpG
SemEval 2007 Task 2	Enough with the reality TV. One broadcaster had an original idea and dumped it. Terra Nova deserves a 2nd chance. Up for it?
SemEval 2007 Task 2	Everything seems to quiet down when USAID was booted out of Russia, may be the money's gone. The "We say when to dissent" movement
SemEval 2007 Task 2	Extremely jealous of Owain's day off tomorrow, I hate Sundays!
SemEval 2007 Task 2	Fast bowler James Pattinson has a buttock injury and won't play in tomorrow's third and deciding one day final against Sri Lanka. #cricket
SemEval 2007 Task 2	FCBarcelona captain, Carles Puyol will be out for two months after dislocating his elbow in the match against SLBenfica on tuesday. #fcblive
SemEval 2007 Task 2	Feeding the dogs an hour later since DST ends soon and I don't want them waking me up at 6am on Sunday. They hate this. I hate DST
SemEval 2007 Task 2	Fellas were giving out Mormon bibles in CIT on Monday, nobody took one. Same on Tuesday. Today they tried to lure us in with sweets..
SemEval 2007 Task 2	Finish taking the Cashee math prep... (-.-) ...Stupid 7th grade math I don't even remember was on it... #Prep
SemEval 2007 Task 2	Fire damages Buddhist temple at SKorea's Mt Naejang: Seoul, Oct 31 (Kyodo) A fire destroyed the main building of... http://t.co/s57mmK5r
SemEval 2007 Task 2	FLASH: Senegal are disqualified from the 2013 Africa Cup of Nations after a riot forced Saturday's home tie with Ivory Coast to be abandoned
SemEval 2007 Task 2	fuck i have the SATs tomorrow
SemEval 2007 Task 2	Fuck the talent show lets go to the basketball tournament on Friday at Richard bland to support kenston forest
SemEval 2007 Task 2	Galaxy S II can be wiped by just clicking a link, other devices may be vulnerable http://t.co/Z643jSTQ #Galaxy #Samsung #HTML #Vulnerability
SemEval 2007 Task 2	Galaxy S II can be wiped by just clicking a link, other TouchWiz devices may also be vulnerable
SemEval 2007 Task 2	Galaxy S II can be wiped by just clicking a link, other TouchWiz devices may also be vulnerable | Th http://t.co/YCezR0qp
SemEval 2007 Task 2	Galaxy S II can be wiped by just clicking a link, other TouchWiz devices may also be vulnerable http://t.co/9Yj0W56z #gaffaw #fb
SemEval 2007 Task 2	Galaxy S II can be wiped by just clicking a link, other TouchWiz devices may also be vulnerable http://t.co/jUK6IzPY #ops
SemEval 2007 Task 2	Gators can't get caught looking ahead to March: So tell me, when does the NCAA Tournament start? http://t.co/2BClV6qR
SemEval 2007 Task 2	Gawd, UAN on Nickelodeon tomorrow. -______-" Remind me to not get home before 6. I hate watching the boys on Nick, it just gets to me. >:(
SemEval 2007 Task 2	Getting sick right before the UCLA football game on Saturday <<<
SemEval 2007 Task 2	got science PSSA's tomorrow. what's the point of retaking them when i did all the study island remediation already? school sucks man.
SemEval 2007 Task 2	Gresh and Zo discussed the Celtics season opener loss to the Heat on Tuesday night. Rajon http://t.co/sbo3Q3Ko
SemEval 2007 Task 2	Gutted I'm missing the cardigan match on Saturday! But more important things to do
SemEval 2007 Task 2	Halloween wore me out next on the list Danksgiving on Nov 23 at Big Fish Pub in tempe hit up for slots and info only a couple left
SemEval 2007 Task 2	Hard day looking forward to my tea now may have to call in to an irlam takeaway #memories
SemEval 2007 Task 2	He going to the fair , on Saturday & I'm going to be out of town ! Mann BULLSHIT
SemEval 2007 Task 2	Heads up: Galaxy S II can be wiped by just clicking a link, other TouchWiz devices may also be vulnerable - http://t.co/1qoMUT8a /r
SemEval 2007 Task 2	Heads up: Galaxy S II can be wiped by just clicking a link, other TouchWiz devices may also be vulnerable http://t.co/hHesGne2
SemEval 2007 Task 2	Heather is like the serpent in the garden trying 2 entice info from McBain! Michael Easton, dont ever go away that long! Watching Tues #GH.
SemEval 2007 Task 2	Hey, the 2nd worst baseball movie is on Reelz right now. #StomachBleeding

SemEval 2007 Task 2	How did Dylan Quirk not make the Vic 19's. He kept to James Pattinson, Darren Pattinson and Peter Siddle in a 1st grade Grand Final at 15.
SemEval 2007 Task 2	How strange in Tia Sharp's case and now April the police have a 'man helping' with inquiries yet they can't conclude what has\is happenec
SemEval 2007 Task 2	http://t.co/N0ynB9vr, Tudou merger may be in jeopardy, report says, Bloomberg notes: See the rest of the story he... http://t.co/wan8PCEx
SemEval 2007 Task 2	http://t.co/zB0dvlFP - This is why the BNP will never get elected. They may put on suits but what is behind them won't be hidden, thank god
SemEval 2007 Task 2	I always thought Michael Dyer & Isiah Crowell would play a big game in Nov for a conference title. Just didn't know it would be the SWAC.
SemEval 2007 Task 2	i am not in the state of mind to take the ACT tomorrow
SemEval 2007 Task 2	I am telling you come what may, Kareena Kapoor will not win a National Award for Heroine.
SemEval 2007 Task 2	I bought an ecig through the sun paper yet not allowed to use it while watching football at Newcastle Utd
SemEval 2007 Task 2	i can't find my school uniform, guess i'm going in naked tomorrow
SemEval 2007 Task 2	I don't even feel like going to LCC on Saturday.
SemEval 2007 Task 2	I don't know if there's a greater indictment of the current GOP than it violating Reagan's 11th Commandment to attack Christie. What a joke.
SemEval 2007 Task 2	I don't like testing I'm glad I don't have to worry about the CASHEE anymore on may.8th (: passed that shxt !
SemEval 2007 Task 2	I don't think I'm going to JOUVERT tomorrow.
SemEval 2007 Task 2	I don't want Saturday to get here. But I'm ready to bomb #allregion in Hot Springs, sooo... Might as well get it over with.
SemEval 2007 Task 2	I get it now he told @clbrooks_48 tomorrow so she wouldn't come. Dang nobody likes Calle.
SemEval 2007 Task 2	I get to work the Bucks game on Thurs now. Im gonna fake high five John Henson and say "Can i borrow your arms to floss my teeth, douche?"
SemEval 2007 Task 2	I got the practice CAHSEE tomorrow at school. dear practice CAHSEE, can you not. sincerely, me.
SemEval 2007 Task 2	I got tickets to the NC State game saturday and nobody to go with..
SemEval 2007 Task 2	I hate when ppl do rates on IG & lie! The bitch look like UNDERTAKER & you gone give her a 9.7 , c'mon man!
SemEval 2007 Task 2	I have a 150 minute flight tomorrow morning. #fuck #sleeping #tonight. I'm going to sleep on the plane, and listen to Yoko Kanno.
SemEval 2007 Task 2	I have HSPA tomorrow why the fuck am I awake ?
SemEval 2007 Task 2	I haven't been able to enjoy tea since Tuesday. I just want to drink a cup without feeling as though I'll hurl
SemEval 2007 Task 2	I just don't know if The Celtics have a chance this season.... also concerned that Garnett may need a hug.
SemEval 2007 Task 2	I just hope that Sandy goes away before November 6 because if she doesn't then VIPs in NJ probably have to cancel the NJ concert :c
SemEval 2007 Task 2	I like Trey Burke but I think pre-season 1st team all-american is a bit over the top.
SemEval 2007 Task 2	I may be an asshole for insulting 1D fans but you guys crossed the line by saying Mitch deserved to die. Stereotypical 1D fans. -_-
SemEval 2007 Task 2	I may end up just watching more NHK tonight. Feel too nauseous to do any actual gaming before bed. Have to get up at 5 AM tomorrow.
SemEval 2007 Task 2	I may or may not have freightened a large number of L.A. residents, merely by making EYE contact with them in the past 48 hours.
SemEval 2007 Task 2	I mean I have to take the OGT tomorrow so why not stay up all night?
SemEval 2007 Task 2	I mean of course like millions of Lovatics will vote for Jillian. She may not know it but she's basically sucking up.
SemEval 2007 Task 2	I miss old school Dashboard Confessional, taking back Sunday, senses fail, early November, and the used. #middleschooljams
SemEval 2007 Task 2	I really wish I didn't have to take the ACT tomorrow
SemEval 2007 Task 2	I think i may be wasting my time with them...So i will just leave it alone but what comes around goes around...Karm is a BITCH
SemEval 2007 Task 2	I think the Sun should consider suing the BNP over this - http://t.co/9ZXqbTE4
SemEval 2007 Task 2	I think those that care about Newcastle Utd. really need to stop this happening. http://t.co/59srzVYL
SemEval 2007 Task 2	I think we may have just found out the reason he plays Carrick at centre half when we have no one else #lackofexperienceshows
SemEval 2007 Task 2	I want to go to the SMB tomorrow but worried i will oversleep :/ been sleeping through my alarm
SemEval 2007 Task 2	I was yelling at Siri cause she wouldn't understand what I was asking her so she played "There's a Great Big Beautiful Tomorrow" from WDW XD
SemEval 2007 Task 2	I would rather eat my left foot then to be taking the SATs tomorrow
SemEval 2007 Task 2	I'll blow up soon & not even remember a day before last june, june 14th, Club Paradise, to be correct.. The last day every thing seemed ok

SemEval 2007 Task 2	I'm absolutely skint til payday next Friday. This all stems from being paid the day I went to Leeds festival.
SemEval 2007 Task 2	I'm going to claytons tomorrow morning but I'm too tired to think about what I'm gonna wear. Looks like another ugly day with him. Oh well
SemEval 2007 Task 2	I'm really sad I'm not going to able to celebrate the 100th Anniversary of the Montreal Canadiens for the 4th year in a row. #NHLLockout
SemEval 2007 Task 2	I'm so sad. No Inkigayo tomorrow. That means we don't get to see IU in her Halloween costume on TV. Hope she tweets a selca. Witch IU?
SemEval 2007 Task 2	i'm so tight because i'm taking the HSPA tomorrow and am gonna miss some of my classes.
SemEval 2007 Task 2	I've narrowed it down to Sophos and Chrome crapping out on chrome!SetPrinterInfo+0xcfc. Detours may make this very difficult to figure out.
SemEval 2007 Task 2	If I feel even worse tomorrow for my chem exam, I'm going to be sooooo MAD
SemEval 2007 Task 2	If I wasn't going to Owain's tomorrow , I would probably spend the whole night crying to myself $: ..
SemEval 2007 Task 2	If Petraeus is ruined then so is McChrystal -he probably handed out the guns in Libya - bc he owes Obama/Biden for his 4th STAR
SemEval 2007 Task 2	If POTUS Barack Hussein Obama during tomorrow nights debate would call Willard Mitt Romney a liar the 1st time Mitt lies, Election is his.
SemEval 2007 Task 2	If Spurs sell VDV, plus Modric leaves, they may as well save us the wait n just join the Championship.
SemEval 2007 Task 2	In tears after watching that Dance Academy episode for the 2nd time
SemEval 2007 Task 2	Intimate showcase with Chris Rene - November 7 in Auckland. Can't buy tickets to this. Click the link to enter... http://t.co/jfhKiRzM
SemEval 2007 Task 2	it doesn't matter that iona doesn't have school tomorrow because 1. i hate school 2. we don't have snow days
SemEval 2007 Task 2	It is too early to be talkin about throwin that ass at NiteMoves on a Sunday Morning take yall hethen ass to Church
SemEval 2007 Task 2	It was windy on Saturday? Andre's fault. Joseph Kony hasn't been caught? Andre's fault. Dell make shit PC's? Andre's fault. #AVB #AVBout
SemEval 2007 Task 2	To the 20 years old Chunji! Happy Birthday, stay healthy and may everything goes as you wish. OH YEAH YOU WISH EVERYTHING GOES YOUR WAY XD
Ukraine Corpus	@StefanFuleEU Ahoj, I'm one of thousands Ukrainians, who standing on #euromaidan now.And I want to EU. Every Ukrainian wants.Please, help us
Ukraine Corpus	Eugenia Tymoshenko tells @France24_en a change of heart by #Ukraine president is her mother's only hope of release for medical treatment
Ukraine Corpus	Welp @JohnKerry has spoken,guess #Ukraine is up next for a "democratic" coup,look for @hrw to be dispatched to get "evidence" of evil #Kiev
Ukraine Corpus	Alarming events in #Ukraine: @globolitics http://t.co/sY961pg8gn
Ukraine Corpus	Clearly #Putin is a complete criminal. #extortion #fraud #theft.....at crazy levels! #winterolympics
Ukraine Corpus	Situation in #Zaporozhye┬áis very worrying. People got injured & avoid hospitals because they fear thugs. #euromaidan http://t.co/Myhb1aylck
Ukraine Corpus	Zakharchenko promised to not arrest people if #euromaidan leaves Hrushevskogo. Better protect mandible, Liar.
Ukraine Corpus	Horrible & shocking conflict, my thoughts & prayers for the people of #Ukraine #Kyiv #Kiev #Support4Ukraine #euromaidan #humanrights
Ukraine Corpus	That's a bloody, not black Friday indeed in #Ukraine! #Euromaidan
Ukraine Corpus	@Pontifex Holy Father please pray for Ukraine today! there is the blood of young citizens on the Main Square of Kyiv! #Euromaidan
Ukraine Corpus	#NTV: #Lukwago & #Besigye siege continues; protests in #Thailand & #Ukraine. Gap in class of protests. Uganda is bigger than two hooligans.
Ukraine Corpus	Following the events in #Ukraine and praying for my friends over there. #euromaidan
Ukraine Corpus	This is rubbish left by a family outside the NN hospital earlier. Useless scanky pieces of worthless shites
Ukraine Corpus	what? he peed on the floor?! #wtf
Ukraine Corpus	Neglecting a child and abusing them is easily one of the lowest things in the world and reserves you a special spot right in hel
Ukraine Corpus	#MSNBC has become a joke - a mirror image of #Fox #Shameful lies & misinfo from #PhilGriffin & his party of ass kissers for a buck
Ukraine Corpus	If you don't like what I have to say about Obama, BLOCK ME! My FATHER took a permanent dirt nap for this turned-to-shit country!
Ukraine Corpus	This was tight slap on indian legal system....Rapist said u didn't hang delhi rapist bt we would hang rape victim #Pathetic #Upgovt
Ukraine Corpus	#AccheDin @dna: Another Dalit girl gang-raped in Uttar Pradesh http://dnai.in/ceh9 "
Ukraine Corpus	Garbage calls. Call it both ways at least
Ukraine Corpus	@peoplemag really makes you wonder what the fuck is wrong with everyone? When did #boys grow #vaginas and become giant #pussies?
Ukraine Corpus	Hill of Tara vandalised - why oh why would someone do this http://ln.is/www.independent.ie/i/nzVq4 ÔÇª @Independent_ie #irelandsheritage
Ukraine Corpus	Celebs think they're so untouchable, drugs are drugs. Weed can turn to harder drugs and then you risk your life. #drugsareformugs

Ukraine Corpus	@iA7med80 Unfortunately, similar cases I know of in #MENA -children accused of looking too sexy & drove men to it! Early Puberty.
Ukraine Corpus	Yuppp the world is coming to an end with the tweet i just saw #eww
Ukraine Corpus	This #Nation #Disgust me sometime i Fully #Respect The #Fact #JasonCollins came out but people are so #Ignorant @jasoncollins34 its #Tragic
Ukraine Corpus	I'm having a battle with ants! -ƒÿ¡-ƒÆÇ-ƒÉ£ Don't know where they're all coming from -ƒÉ£-ƒÉ£-ƒÉ£-ƒÉ£-ƒÉ£-ƒÆó
#disgust	ÔÇ£@Dr_JamesLogan: I'm mostly disgusted by feet and spiders. #cheltscifestÔÇØ eeewww feet #blah
#disgust	WHYYYY is my finger skin peeling.... wtf. -___-
#disgust	I forgot how awful it smells by the trash compactor
#disgust	Joser is so fckn ugly, always calls my family "blazdashians"
#disgust	My car is green from all this pollen
#disgust	@Michael_Corazza you're the number one scavenger pic.twitter.com/j9dbtUTEmc
#disgust	Kim Kardashians dress actually looked gross
#disgust	Seriously there are people who make me sick. I question their morals, values, & if they even have a brain?!?#somepeoplemakemesick
#disgust	I really don't understand how my brother is so skinny when all he eats is junk food. #twentyfivecookies #ew
#disgust	Do guys really think licking their lips at a girl works? Because it most certainly does not
#disgust	Think it's absolutely foul how some parents actually want their child to have a disability. Probably to get the easy life on benefits!
#disgust	he couple in front of me in the parking lot just straight up started licking each other...
#disgust	Wanted to save my prom flowers cause they were pretty. I didn't know flowers would mold ... pic.twitter.com/KWohIQlDAE
#disgust	That's no reason to cut off your skin that'd be way too much to clean up! pic.twitter.com/6vFR3o62DI
#disgust	Remember that time I jammed my finger at practice :-) pic.twitter.com/g7pnY6Z6H6
#disgust	I wish some mothers of trashy girls really knew how nasty their daughters are an would do a better job raising them
#surprise	Had the best night with @LouiseSammi surprising this beaut for her 21st-ƒÆ£-ƒÄé-ƒÄê-ƒí╢ @sophie_hanspal #birthday #21 pic.twitter.com/mcHANVDDau
#surprise	Gotta love when you are randomly thrown into a 4x800m relay when you have never ran more than a 400.
#surprise	Sneaked into the bae's house and she has no idea #naptime
#surprise	Was so into Yeezus that I didn't realize the gym was closing and I was the only person there
#surprise	Boyfriend discovers that his 'pregnant' girlfriend is actually a man http://us.tomonews.net/89454809038848
#surprise	Left Lex's number on the table at the restaurant tonight and the waiter texted her
#surprise	Just ordered 5 chickens online without my mom knowing. I wonder what she's guna say when they arrive in the mail on May 28th

Bibliography

[1] Francis Joseph Aguilar. *Scanning the business environment*. Macmillan New York, 1967.

[2] A. Alaszewski. *Using Diaries for Social Research*. Introducing Qualitative Methods series. SAGE Publications, 2006.

[3] Effie Amanatidou, Maurits Butter, Vicente Carabias, Totti Könnölä, Miriam Leis, Ozcan Saritas, Petra Schaper-Rinkel, and Victor van Rij. On concepts and methods in horizon scanning: Lessons from initiating policy dialogues on emerging issues. *Science and Public Policy*, 39(2):208–221, 2012.

[4] Shenghua Bao, Shengliang Xu, Li Zhang, Rong Yan, Zhong Su, Dingyi Han, and Yong Yu. Mining social emotions from affective text. *IEEE Trans. Knowl. Data Eng.*, 24(9):1658–1670, 2012.

[5] Erik Cambria, Marco Grassi, Amir Hussain, and Catherine Havasi. Sentic computing for social media marketing. *Multimedia Tools Appl.*, 59(2):557–577, 2012.

[6] Erik Cambria, Catherine Havasi, and Amir Hussain. Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis. In G. Michael Youngblood and Philip M. McCarthy, editors, *FLAIRS Conference*. AAAI Press, 2012.

[7] Erik Cambria and Amir Hussain. *Sentic computing*. Springer, 2012.

[8] Erik Cambria, Andrew Livingstone, and Amir Hussain. The hourglass of emotions. In *Cognitive behavioural systems*, pages 144–157. Springer Berlin Heidelberg, 2012.

[9] Erik Cambria, Bjorn Schuller, Yunqing Xia, and Catherine Havasi. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, page 1, 2013.

[10] Erik Cambria, Robert Speer, Catherine Havasi, and Amir Hussain. Senticnet: A publicly available semantic resource for opinion mining. In *AAAI Fall Symposium: Commonsense Knowledge*, volume FS-10-02 of *AAAI Technical Report*. AAAI, 2010.

[11] François-Régis Chaumartin. Upar7: A knowledge-based system for headline sentiment tagging. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 422–425. Association for Computational Linguistics, 2007.

[12] Jorge Carrillo de Albornoz, Laura Plaza, and Pablo Gervás. Sentisense: An easily scalable concept-based affective lexicon for sentiment analysis. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet U?ur Do?an, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may 2012. European Language Resources Association (ELRA).

[13] Marie-Catherine De Marneffe, Bill MacCartney, Christopher D Manning, et al. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, volume 6, pages 449–454, 2006.

[14] Food DEFRA, Department for Environment and Rural Affairs. Horizon scanning, March 2009.

[15] Paul Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3-4):169–200, 1992.

[16] Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining, 2006.

[17] Christiane Fellbaum, editor. *WordNet: an electronic lexical database*. MIT Press, 1998.

[18] Virginia Francisco, Pablo Gervás, and Federico Peinado. Ontological reasoning for improving the treatment of emotions in text. *Knowl. Inf. Syst.*, 25(3):421–443, 2010.

[19] Dennis Grinberg, John Lafferty, and Daniel Sleator. A robust parsing algorithm for link grammars. *arXiv preprint cmp-lg/9508003*, 1995.

[20] Beat Habegger. Horizon scanning in government: Concept, country experiences, and models for switzerland. 2009.

[21] Alan R. Hevner. Design science in information systems research. *MIS quaterly*, 28(1):75–105, 3 2004.

[22] Klerx Joachim, Mak Klaus, and Göllner Johannes. Foresight in der strategischen planung. to appear.

[23] Klerx Joachim, Mak Klaus, and Göllner Johannes. Management of horizon scanning centers – foresight and horizon scanning. to appear.

[24] David A King and Sandy M Thomas. Taking science out of the box-foresight recast. *SCIENCE-NEW YORK THEN WASHINGTON-*, 316(5832):1701, 2007.

[25] Satoshi Morinaga, Kenji Yamanishi, Kenji Tateishi, and Toshikazu Fukushima. Mining product reputations on the web. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, pages 341–349, New York, NY, USA, 2002. ACM.

[26] Preslav Nakov, Zornitsa Kozareva, Alan Ritter, Sara Rosenthal, Veselin Stoyanov, and Theresa Wilson. Semeval-2013 task 2: Sentiment analysis in twitter. 2013.

[27] Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. Textual affect sensing for sociable and expressive online communication. In Ana Paiva, Rui Prada, and Rosalind W. Picard, editors, *ACII*, volume 4738 of *Lecture Notes in Computer Science*, pages 218–229. Springer, 2007.

[28] Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. Emoheart: Conveying emotions in second life based on affect sensing from text. *Adv. Human-Computer Interaction*, 2010, 2010.

[29] Greg J Norman, Catherine J Norris, Jackie Gollan, Tiffany A Ito, Louise C Hawkley, Jeff T Larsen, John T Cacioppo, and Gary G Berntson. Current emotion research in psychophysiology: The neurobiology of evaluative bivalence. *Emotion Review*, 3(3):349–359, 2011.

[30] Alexander Osherenko. *Opinion mining and lexical affect sensing.* PhD thesis, University of Augsburg, 2010. http://d-nb.info/1009392654.

[31] Olutobi Owoputi, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A. Smith. Improved part-of-speech tagging for online conversational text with word clusters. In *In Proceedings of NAACL*, 2013.

[32] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.

[33] James W Pennebaker, Cindy K Chung, Molly Ireland, Amy Gonzales, and Roger J Booth. The development and psychometric properties of liwc2007. *Austin, TX, LIWC. Net*, 2007.

[34] James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71:2001, 2001.

[35] Robert Plutchik. The nature of emotions human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4):344–350, 2001.

[36] Soujanya Poria, Basant Agarwal, Alexander Gelbukh, Amir Hussain, and Newton Howard. Dependency-based semantic parsing for concept-level text analysis. In *Computational Linguistics and Intelligent Text Processing*, pages 113–127. Springer, 2014.

[37] Soujanya Poria, Alexander F. Gelbukh, Amir Hussain, Newton Howard, Dipankar Das, and Sivaji Bandyopadhyay. Enhanced senticnet with affective labels for concept-based opinion mining. *IEEE Intelligent Systems*, 28(2):31–38, 2013.

[38] Dheeraj Rajagopal, Erik Cambria, Daniel Olsher, and Kenneth Kwok. A graph-based approach to commonsense concept extraction and semantic similarity detection. In *Proceedings of the 22nd international conference on World Wide Web companion*, pages 565–570. International World Wide Web Conferences Steering Committee, 2013.

[39] James A. Russell. Core affect and the psychological construction of emotion. *Psychological Review*, pages 145–172.

[40] James A Russell, Anna Weiss, and Gerald A Mendelsohn. Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology*, 57(3):493, 1989.

[41] Philip J. Stone. *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press, 1966.

[42] Carlo Strapparava and Rada Mihalcea. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 70–74. Association for Computational Linguistics, 2007.

[43] Carlo Strapparava and Alessandro Valitutti. Wordnet affect: an affective extension of wordnet. In *LREC*. European Language Resources Association, 2004.

[44] William J Sutherland, Mick Clout, Isabelle M Côté, Peter Daszak, Michael H Depledge, Liz Fellman, Erica Fleishman, Rachel Garthwaite, David W Gibbons, Jennifer De Lurio, et al. A horizon scan of global conservation issues for 2010. *Trends in Ecology & Evolution*, 25(1):1–7, 2010.

[45] Yla R. Tausczik and James W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. 2010.

[46] Mike Thelwall. Heart and soul: Sentiment strength detection in the social web with sentistrength. *Cyberemotions*, pages 1–14, 2013.

[47] Yoshimasa Tsuruoka and Jun'ichi Tsujii. Bidirectional inference with the easiest-first strategy for tagging sequence data. In *Proceedings of HLT/EMNLP 2005*, pages 467–474, 2005.

[48] Inc Twitter. About twitter, April 2014.

[49] Mohamed Yassine and Hazem Hajj. A framework for emotion mining from text in online social networks. In Wei Fan, Wynne Hsu, Geoffrey I. Webb, Bing Liu, Chengqi Zhang, Dimitrios Gunopulos, and Xindong Wu, editors, *ICDM Workshops*, pages 1136–1142. IEEE Computer Society, 2010.