A Generic Approach to Generate Opinion Lists of Phrases for Opinion Mining Applications

Sven Rill
University of Applied Sciences
Alfons-Goppel-Platz 1
95028 Hof, Germany
srill@iisys.de

Johannes Drescher
University of Applied Sciences
Alfons-Goppel-Platz 1
95028 Hof, Germany
jdrescher@iisys.de

Dirk Reinel
University of Applied Sciences
Alfons-Goppel-Platz 1
95028 Hof, Germany
dreinel@iisys.de

Jörg Scheidt
University of Applied Sciences
Alfons-Goppel-Platz 1
95028 Hof, Germany
jscheidt@iisys.de

Oliver Schütz
University of Applied Sciences
Alfons-Goppel-Platz 1
95028 Hof, Germany
oliver.schuetz@hof-university.de

Florian Wogenstein
University of Applied Sciences
Alfons-Goppel-Platz 1
95028 Hof, Germany
fwogenstein@iisys.de

ABSTRACT
In this paper we present an approach to generate lists of opinion bearing phrases with their opinion values in a continuous range between -1 and 1. Opinion phrases that are considered include single adjectives as well as adjective-based phrases with an arbitrary number of words. The opinion values are derived from user review titles and star ratings, as both can be regarded as summaries of the user's opinion about the product under review. Phrases are organized in trees with the opinion bearing adjective as tree root. For trees with missing branches, opinion values then can be calculated using trees with similar branches but different roots. An example list is produced and compared to existing opinion lists.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering*; I.2.7 [**Natural Language Processing**]: Text analysis

General Terms
Algorithms

Keywords
Opinion bearing phrases, opinion mining, text resource

1. INTRODUCTION

1.1 Lexical Resources for Opinion Mining
As much as the amount of textual data increases in the world wide web, the need of techniques for automatically analyzing this data gains weight. Especially for analyzing user generated content, opinion mining came into the focus of many research activities.

Many algorithms for the automatic extraction of opinions from textual data need text resources like polarity lexicons consisting of words or phrases with their opinion values.

These opinion values either just indicate a positive, neutral or negative polarity with discrete values (-1, 0 and $+1$) or take continuous values between -1 and $+1$ providing a finer resolution in the measure of their opinion polarities.

Polarity lexicons can be produced manually or derived from dictionaries or text corpora. As there are at least several hundreds of commonly used opinion bearing words in a wide spread language, a pure manual collection and rating of them is not applicable. The dictionary based approach has the disadvantage that it does not take the fact into account that the polarity of many opinion words vary depending on the context they are used in. Deriving opinion values from a text corpus may make it possible to deal with this problem. In [17] the aforementioned three approaches are discussed in detail.

The quality of the lexical resources used is of utmost importance for the quality of the results obtained from opinion mining applications. In [25] the authors applied opinion mining on customer feedback data. They analyzed error sources and came to the conclusion that about 20% of the errors occurred due to faults in the opinion list. In addition, most of the other error sources were related to the opinion list.

1.2 Opinion Words and Phrases
Hereinafter, we want to explain why we consider it helpful to include opinion values for phrases in our opinion list.

Having an opinion value for one single word, one problem that often occurs is to handle phrases containing the opinion word plus one or several valence shifters (intensifiers or reducers) or negation words as well as combinations of both.

One could assume that negation words just change the sign of the opinion value and valence shifters change its absolute value by a defined step. In some cases, however, this is not correct. For example, "good" and "perfect" are positive opinion words, "not good" can be regarded as a negative phrase, although "not perfect" cannot.
Similar effects occur for valence shifter words. The intensifier word "very" does not change the opinion value of different words by exactly the same amount. The amount of the shift rather depends on the distance to the maximum possible opinion value (± 1).
In principle, there are two possible approaches to handle these problems:

- The first one is a sophisticated treatment of intensifiers, reducers and negation words during the application of a list providing opinion values for single opinion words plus values for valence shifters and negation words. This sentiment composition is discussed in several publications [5, 14, 18, 21].

- The second possibility is to provide a list with opinion values for phrases including intensifiers, reducers and negation words.

Another big issue in the field of lexical-based opinion mining is the problem of domain- and aspect-specific opinion words. Simple opinion words like "good", "bad" and "great" have unambiguous opinion polarities whereas others do not. A drastic example for a domain-specific difference in opinion values is the word "scary". Having a quite negative meaning in most of the domains, for horror books or films the statement "The book is scary" probably is meant positively. As an example of an aspect-specific difference consider "The camera is small" and "The display is small". The first statement might be meant positively, the second negatively.
One way to handle this problem could be the provision of stable algorithms making it possible to derive domain- and aspect-specific lexical resources quite easily.
We introduce a quite generic approach to derive an opinion list containing opinion bearing phrases together with their opinion values.

The used data base consists of titles and star ratings from user reviews. The idea is to organize opinion bearing words plus shifters and negation words in trees providing opinion values for each tree vertex.

2. RELATED WORK

Especially in the last decade, a lot of research work has been done in the area of opinion mining. A detailed overview of the whole topic recently has been given in [17].
Several publicly available text resources, especially lists of opinion bearing words, have been provided for several languages.
For the English language commonly used resources are SentiWordNet (SWN) [2, 7, 8], Semantic Orientations of Words (SOW) [30], the Subjectivity Lexicon (SL) [35] and two lists of positive and negative opinion words provided by [16].
SWN and SOW were generated using the WordNet® [20] lexical database.
SWN includes almost 118,000 words. About 18,000 of them are adjectives, the others are verbs, adverbs and nouns. It provides the probabilities of expressing a positive or negative opinion or being objective for each word. The probabilities are given in discrete steps of one-eighth and add up to one. In addition, different meanings of words lead to multiple entries with different probabilities. This allows for a context-specific treatment in opinion mining applications.
SOW lists about 88,000 words. About 20,000 of them are adjectives. The opinion bearing words with a strong polarity have opinion values of $+1$ and -1. Weak opinion bearing words are listed with opinion values on a continuous scale in the range of ± 0.4.
SL was generated by expanding a list of subjectivity clues from [29] based on a dictionary. It contains about 8,200 opinion words where 3,200 of them are adjectives. For each word the polarity (positive or negative) is provided together with a strength of the subjectivity (strong or weak). A word is assumed to be a strong subjective word if it is subjective in most of the possible contexts. If a word is only subjective in several contexts, it is marked as weak subjective.
All three lists do neither handle phrases nor provide factors for shifter or negation words. SWN and SL will be the basis for a first evaluation of our opinion list (see Chapter 4).
In [1], a polarity lexicon is derived using a graph propagation framework. The main focus lies on the extraction of slang, urban opinion words and misspellings. A similar approach is followed in [33], where the construction of web-derived opinion lexicons is described. In [36] various aspects of the extraction of opinions from social media data are discussed. These include both the calculation of word scores for adjectives and adverbs based on review texts and the definition of an algorithm to derive sentence scores based on these word scores.
Further lists of opinion values also exist for other languages [3, 6, 28, 34].
Several groups use online reviews for different research projects. Typical research topics are classification [26, 32], summarization [24], spam detection [11], feature extraction [27] and the analysis of the helpfulness of reviews [22]. For an overview of the whole area of review analysis see [31].

3. GENERATION OF THE OPINION LIST

3.1 Idea and Overview

As described in Chapter 1, the aim of this work is to provide an algorithm for building a list containing opinion bearing phrases together with their opinion values on a scale between -1 and $+1$. The algorithm is based on user written product reviews.
These reviews include, among other information, a star rating, the review title and the review text. The title as well as the star rating can be regarded as a summary of the review text. Thus, it is obvious that the opinion expressed via the star rating is strongly correlated to the one expressed with opinion bearing words and phrases in the review title. Due to this, the opinion value for a word or phrase occurring in the titles of reviews can be derived from the star ratings assigned by the review authors.
Basic input for the algorithm are pairs of review titles and star ratings. This implies that the algorithm is applicable for all user review systems providing at least a title and some numerical assessment.
Figure 1 depicts the whole system used to derive the opinion values.

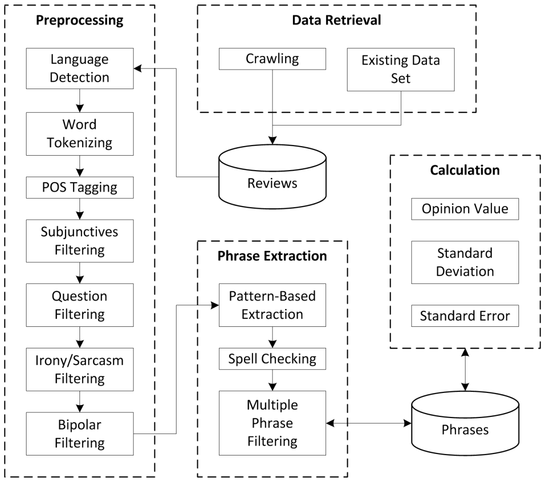

Figure 1: Overview of the opinion list generation.

3.2 Data Retrieval and Preprocessing

3.2.1 Crawling and Language Detection

As a basis we took the publicly available collection of Amazon.com reviews provided by [12] containing about 5.8 million user reviews. This data source has already been used for several research works such as [13, 15, 23]. We enriched it with additional data crawled from the Amazon.com website. After that, the whole data set consisted of about 6.9 million pairs of review titles and star ratings.
A language detection, used to exclude other than English language review titles, was not necessary for the present list. Nevertheless, it is part of the algorithm as it will be necessary for building opinion lists for other languages, e.g., for German, as reviews from the Amazon.de website also contain some fractions of English reviews.

3.2.2 Word Tokenizing and Part-of-Speech Tagging

The next preprocessing steps to perform are the word tokenizing and the part-of-speech (POS) tagging. The POS tagging is a crucial point for this project due to following reasons:

- A correct POS tagging is essential because the phrase construction should be restricted to adjectives. Including faulty (as adjective) tagged nouns or verbs will result in wrong contributions for the opinion list.
- Adjectives wrongly tagged as nouns will not falsify the result but will be excluded from the calculation of the opinion values.

On the other hand, POS tagging is more difficult for review titles as they often are not complete sentences but sometimes consist only of some words (e.g., "Great Mobile Phone") and, therefore, are quite often mistagged. In particular if an adjective is the first word in the title and thus starting with a capital letter, it is often mistagged as a noun or named entity, e.g., in "Smart story", "smart" has to be tagged as an adjective. We used the Apache OpenNLP POS Tagger[1], with the maximum entropy model which was trained using the Penn Treebank Tagset [19].
To be able to quantify the error of faulty POS tags for the review titles, we manually tagged about 220 review titles and compared these results with the results of the automated POS tagging. We obtained a precision value of $p = 0.92$ and a recall value of only $r = 0.64$ indicating that many adjectives are not tagged as such.
As we stated before, the main error source is the fact that words in capital letters, e.g., at the beginning of the review, are often mistagged. To improve the POS tagging, we converted all words in the titles to small letters and repeated the POS tagging. Repeating the check using the manually tagged sample, it turns out that the precision value stays at $p = 0.92$ while the recall value improves to $r = 0.94$.

3.2.3 Filtering

As explained in Section 3.1, the main idea of our approach is that a star rating corresponds to the polarity of an opinion phrase in a review title. As the opinion value for the phrases will be calculated on this basis in a later step, titles for which this assumption is not true have to be discarded.
Therefore, titles are filtered out if one of the following cases occur:

- Subjunctives often imply that a statement in a review title is not meant as the polarity of the adjectives indicates, e.g., "Could have been a good film" Therefore, review titles with subjunctives are discarded. In the English language "could", "should" and "would" are typical words indicating subjunctives. Hence, review titles with one of these words are omitted.
- Some titles are formulated as questions. Many of these titles are not useful as adjectives included often express the opposite opinion compared to the star rating, e.g., "Why do people say that this is good?". Therefore, titles are excluded if they contain an interrogative and a question mark at the end.
- Some review titles are meant ironically. Irony cannot be detected automatically in many cases [4] but exceptions exist. Sometimes, for example, writers finalize the title with an emoticon like ";-)". Others put ironically meant words in quotation marks, e.g., "Really a 'great' movie!". Both can be regarded as signs of irony. These titles are excluded from the data set.
- The word "but" is an indicator for a bipolar opinion, e.g., "seems good but ...". Again, the star rating does not correspond to the opinion value of the adjective in the title. For this reason, titles containing a "but" are omitted.

3.3 Opinion Phrase Extraction

The next step is the extraction and filtering of opinion phrases. Opinion phrases consist of at least one opinion bearing word. In addition, they might contain shifters and/or negation words and/or other words like adverbs or other adjectives.
We distinguish between adjective-based phrases and others. Since adjectives are the most common part of speech to express opinions, we just regard adjective-based opinion phrases at the moment. Typical examples for adjective-based phrases are "absolute brilliant", "not very good" and

[1]http://opennlp.apache.org/

"excellent". Examples for the others are "complete rubbish" or "never disappoints (me)".
Now we describe how adjective-based phrases are extracted. Every word tagged as an adjective is a candidate for an opinion phrase. We start the phrase extraction from the end of the title. For each candidate the phrase is extended to the left as long as it fulfills one of the patterns below. As mentioned in Section 3.2.2, the POS tagger returned the Penn Treebank POS tags so the patterns are given using these tags.

1. Single adjective, e.g., "Great!", "Better than expected", "The best notebook!" (JJ[2])
2. One or more adverbs (or their comparative or superlative form) and an adjective, e.g., "Very good film", "A very good watch", "Not good for iPad"[3] (RB+JJ).
3. One or more adverbs, a determiner (as a, an, that) and an adjective, e.g., "Not a good DVD" and "Not that good". Note that the adverb in front of the determiner is mandatory, "a good" is not a phrase allowed here (RB+DT+JJ).
4. Patterns 2 and 3 but with one or more adverbs replaced by adjectives, e.g., "Very nice little screen!".

At this stage of the algorithm, a spell checker is applied to identify misspelled words in the phrases. We used the Hunspell Spell Checker[4] with the en_US wordlist[5]. In cases where the spell checker marks a word as misspelled, this review title is omitted.
In addition, only titles with exactly one adjective-based opinion phrase are accepted at this point. The reason is that titles containing more than one opinion phrase have the problem that phrases normally have different opinion values. In extreme cases they are contradicting, e.g., "Good songs, bad sound quality". Therefore, titles having two or more opinion phrases are discarded.

3.4 Calculation of Opinion Values

After the preselection steps described in the sections before, the data set consists of about 2.9 million review titles each having one adjective-based opinion phrase and a star rating between one and five. For each phrase i occurring frequently in the review titles, the mean star rating SR_i is calculated from the star rating S_j^i of all n review titles having this phrase. Frequently at this stage means that a phrase has to occur at least ten times in the preselected review titles.
Afterwards, the opinion value OV_i is obtained by transposing the mean star rating to the scale $[-1, +1]$:

$$SR_i = \frac{\sum_{j=1}^{n} S_j^i}{n} \qquad OV_i = \frac{SR_i - 3}{2} \tag{1}$$

Please note that in this step we assume that a three star rating represents a neutral rating. In Section 4.5 we will discuss whether this assumption can be regarded as being adequate.

[2]JJ stands for JJ, JJR and JJS, so the comparative and superlative forms are also allowed.
[3]The negation word "not" is also tagged as an adverb.
[4]http://hunspell.sourceforge.net/
[5]http://wordlist.sourceforge.net/

In addition to the opinion value, two quality measures are calculated. The first one is just the standard deviation σ_{OV} of the opinion value. It is a measure of how much the star rating spreads for a given opinion phrase. The second one is the standard error calculated by dividing the standard deviation by the square root of the number n_i of review titles having phrase i. In addition to the spread of the stars, it indicates on how many review titles the opinion value of a given phrase is based.
Table 1 shows some adjective-based phrases together with their opinion values and the two quality measures.

Phrase	**OV**	$\boldsymbol{\sigma_{OV}}$	$\boldsymbol{SE_{OV}}$
absolutely fantastic	0.984	0.091	0.003
great	0.846	0.324	0.001
very lovely	0.783	0.284	0.052
very bad	-0.830	0.421	0.015
just awful	-0.939	0.235	0.015

Table 1: Some adjectives and phrases with their opinion values and two quality measures.

4. RESULTS AND DISCUSSION

4.1 Statistical Summary

Our list consists of 9,012 opinion bearing words and phrases. Each of them is based upon at least ten occurrences in review titles. This raw list contains 3,816 single opinion words as well as 4,456 two-word, 661 three-word and 76 four-word phrases plus three phrases with more than four words. A first look at the list shows that it should be cleaned as it still contains unwanted words and phrases, especially due to mistakes during the POS tagging. A simple way of first cleaning the list is to require a more frequent occurrence in the review titles. It is also clear that the longer the phrases, the less often they are used in titles. This leads to the conclusion that the frequency required should be chosen depending on the length of the phrases.
Table 2 summarizes the length of the opinion list depending on the threshold on the number of occurrences of the opinion words and phrases in the review titles.

	Phrases' Length				
	1	**2**	**3**	**4+**	**Total**
Threshold	10	10	10	10	
No. of Phrases	3,816	4,456	661	79	**9,012**
Threshold	100	50	20	15	
No. of Phrases	1,198	1,027	331	49	**2,605**
Threshold	500	100	50	20	
No. of Phrases	451	516	141	30	**1,138**

Table 2: Number of opinion phrases depending on the threshold on the frequency of the occurrence in the review titles.

Taking the thresholds of the last scenario and getting a list with 1,138 opinion words and phrases, one can be interested in the number of phrases bearing a strong opinion (e.g., with $|OV| \geq 0.7$). We find that in this list 650 (57%) phrases are strong opinion words and phrases. A huge majority of these words (about 600) are positive opinion words. One reason

is that the shapes of the distributions of online product reviews are typically asymmetric (sometimes called “J-shaped distribution”). This has been discussed in several works [9, 10]. Hence, positive words occur more often in review titles and therefore the threshold on the frequency of occurrence is passed more frequently. So, for the final list asymmetric thresholds should be chosen for positive and negative words and phrases.
Another reason is the shifted zero-line of the opinion values (see Section 4.5).

4.2 Opinion Values for Single Adjectives

Table 3 shows some examples of opinion values at this stage of the process.

Adjective	OV	Adjective	OV
awesome	0.949	fraudulent	-0.908
superb	0.947	worthless	-0.899
brilliant	0.932	abysmal	-0.879
delightful	0.870	awful	-0.871
helpful	0.643	incompetent	-0.607
decent	0.233	unclear	-0.254

Table 3: Opinion values for single adjectives.

Note that the absolute values seem to be a bit smaller for the negative opinion words compared to the positive ones. One would expect to be $|OV("awesome")| \approx |OV("awful")|$. Possible reasons for this will be discussed in Section 4.5.1.

4.3 Opinion Values for Phrases

4.3.1 Selected Opinion Phrases

In Table 4 examples for both positive and negative phrases with their opinion values are listed.

Phrase	OV	Phrase	OV
unbelievably great	1.000	truly horrible	-1.000
simply fantastic	0.982	unbelievably bad	-0.989
simply awesome	0.978	simply awful	-0.978
still amazing	0.953	just bad	-0.905
pretty cool	0.610	very stupid	-0.640
fairly interesting	0.180	barely adequate	-0.317

Table 4: Frequent opinion phrases of both polarities with their opinion values.

An opinion value of ±1 means that a phrase did only occur in five star or one star reviews respectively. It is interesting to see that no single word did reach a value of ±1. This can be interpreted in the way that no single word is so extremely polarized that it is used exclusively in very good or very bad reviews while some phrases with two or more words are.

4.3.2 Phrases Based on the Word “good"

As an example for both, an opinion bearing adjective and phrases based on this adjective, we want to consider the word “good”. In the text corpus of several million review titles, “good” is a very frequent word and a lot of phrases are based on it. ”good” as a single word has an opinion value of 0.560 and occurs about 130,000 times in sample. The comparative “better” and the superlative “best” have values of 0.584 and 0.907 respectively.
In addition, we find 111 different two-word, 74 three-word and 19 four-word phrases. Table 5 summarizes some frequent phrases based on “good” together with their opinion values and the frequency of their occurrence.

Phrase	OV	Frequency
so good	0.831	2,867
really good	0.798	3,967
very good	0.755	20,396
so far so good	0.719	1,472
good	**0.560**	**129,405**
pretty good	0.442	10,138
not that good	-0.399	473
not very good	-0.599	885
not good	-0.637	2,195

Table 5: Some examples of opinion values for phrases based on the adjective “good” with their frequency of occurrence.

The opinion values look reasonable in the sense that all shifters change the opinion values in the right direction. It is curious that also the results for the multiple usage of an intensifier word lead to reasonable opinion values: for “very good” we obtain 0.755, for “very very good” 0.854 and for “very very very good” 0.905. More about intensifier words will be discussed in Section 4.3.4

4.3.3 Opinion Values for Missing Phrases

In Chapter 1 we already mentioned the problem concerning the universal use of shifter values for intensifiers, reducers and negation words. We came to the conclusion that a general approach of shifter values does not work. Otherwise, one has to handle phrases which could not be extracted from the review titles. As an example, let us look again at phrases based on the adjective “good”. As discussed in the previous Section, the opinion list contains the opinion values of many phrases based on the adjective “good”. We can organize these phrases in a tree with “good” as root. Each vertex opinion value stands for the phrase we get following the path from the vertex to the root. For example the vertex “not | -0.599” stands for the phrase “not very good” and has an opinion value of -0.599 (see Figure 2, only a small excerpt of the tree is shown).

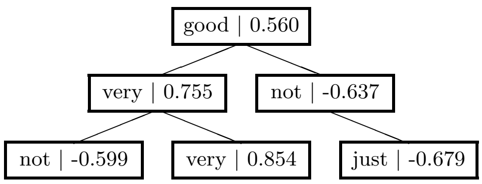

Figure 2: Excerpt from the opinion value tree for the root “good”.

This tree can be used to calculate missing opinion values for other phrase roots.

Let us assume that the tree for a less frequent opinion word looks similar to the tree for "good" in the sense that similar phrases with their opinion values exist. If an opinion value for a phrase that is missing in the list is needed in an application, this value can be derived from the tree of an opinion word with a similar polarity. Let OV_g be the opinion value for "good", OV_{vg} the opinion value for "very good" and so on. Let the less frequent opinion word be "fine", so OV_{vf} stands for the opinion value of "very fine". We assume that the value for "very very fine" is missing in the list and $OV_{vf} = 0.774$.

- The simplest possible approach would be to take the absolute change of the opinion values from the "good"-tree: $OV_{vvf} = OV_{vf} + (OV_{vvg} - OV_{vg}) = 0.873$.
- Another possibility would be to take the relative change of the opinion value compared to the maximum / minimum possible value (± 1). The idea is that in the "good"-tree, the "very" changes the opinion value from 0.755 to 0.854 which is 40% of the possible maximum change (to +1). As "very fine" has an opinion value of 0.774, this approach would lead to an opinion value of: $OV_{vvf} = OV_{vf} + (1 - OV_{vf}) \cdot \frac{OV_{vvg} - OV_{vg}}{1 - OV_{vg}} = 0.865$.

The second approach would lead to a change very similar to the change in the "good"-tree as "very fine" has a similar opinion value as "very good". For a word with a significantly smaller or bigger value, it would result in a bigger / smaller change respectively.

4.3.4 Valence Shifter and Negation Words

As our opinion list contains many phrases with shifters and negation words, the effect of some of these should be analyzed now. We look, e.g., at the shifter words "very" and "pretty" and at the negation word "not".
The word "very" is expected to increase the absolute value of all opinion values. To analyze the effect in detail, we selected all adjectives from our list where the intensified phrase occurred at least 50 times and having an opinion value of ± 0.5 at least.
We find an average value shift $(|OV_P - OV_{vP}|)$[6] of 0.067 but with a standard deviation of 0.166. As expected, the shift value is bigger for adjectives with a smaller opinion value. For example, the opinion value of "good" (0.560) is shifted by 0.195 to 0.755 in the phrase "very good". For the word "great", already having an opinion value of 0.846, the intensifier "very" changes the value only by 0.048 to 0.894.
A little bit different is the word "pretty". Here, we find that for most of the words it decreases the strength of the polarity expressed by a small amount (less than 0.15), e.g., in "pretty good", "pretty cool" and "pretty awful". But in some cases the strength of the opinion polarity is increased by a big amount (more than 0.3), e.g., in "pretty bad" and "pretty sad".
Both findings favor the approach of directly deriving opinion values for phrases instead of attempting to work with generic shifter values.
Table 6 shows the effect of the negation word "not" on several adjectives.

[6] OV_P stands for the opinion value of the phrase itself and OV_{vP} for the intensified phrase.

P	OV_P	$OV_{\overline{P}}$	**P**	OV_P	$OV_{\overline{P}}$
helpful	0.643	-0.711	perfect	0.904	0.294
good	0.560	-0.637	great	0.846	-0.235
friendly	0.508	-0.500	easy	0.755	0.079
bad	-0.288	0.317	terrible	-0.833	-0.063

Table 6: Effect of the negation word "not" on selected opinion words.

It turns out that there a two typical effects of the word "not":

- "not" changes the sign of the opinion value leaving its absolute value nearly unchanged. This is the case for words on the left side of the table.
- "not" changes the polarity from a strong polarity to a weak or zero polarity. Examples for this case are the words on the right side of the table.

The second scenario indicates that for one group of words a generic treatment of negation words is not applicable.

4.4 Comparison to Existing Opinion Lists

In order to compare our list with existing lists, we identify words which are assumed to express a (strong) positive or negative opinion. As the existing lists do not contain opinion values for phrases, only single opinion words are chosen. We compare our opinion values with two benchmark lists explained in Chapter 2. The result of this simple test is summarized in Table 7.
The opinion values of all three lists agree very well.

Word	**OV**	**SWN**	**SL**
extraordinary	0.923	p:0.625; n:0.000	sp
perfect	0.904	p:0.625; n:0.125	sp
good	0.560	p:0.750; n:0.250	wp
lousy	-0.744	p:0.000; n:0.750	sn
useless	-0.812	p:0.125; n:0.625	wn
awful	-0.871	p:0.125; n:0.625	sn

Table 7: Self-generated opinion values (OV) in comparison to SentiWordNet (SWN) and Subjectivity Lexicon (SL).

We regard the inclusion of opinion phrases in our list as an advantage, as the compositional treatment of valence shifters or negation words is quite problematic. In addition, our opinion values are given with a much finer granularity compared to the discrete steps used in the benchmark lists. This might allow a better distinction of opinions in applications using our list of opinion bearing words and phrases.

4.5 Shortcomings and Future Work

4.5.1 Known Problems

At first glance, the results of our approach seem to be quite promising. However, there are several shortcomings which require a closer view leading to the conclusion that some research work remains to be done. Here, we list the known issues emerged during the construction of the opinion list:

- Some opinion phrases appear in idioms expressing a very clear opinion. This sometimes leads to an unexpected high or low opinion value. In this respect, "enough good" has a high opinion value because it is often used in the idiom "Can't say enough good things about ...".

- Especially strong positive words have reasonable opinion values (see Table 3). Unfortunately, for some words with only a weak polarity as well as for words with a strong negative polarity the opinion values seem to be shifted to positive values. The reason for this may be that the assumption of a three star rating representing a neutral opinion (see Equation 1) on a one to five scale may not be true. Some people regard already a three star rating as a somehow negative one, so the zero line is shifted. In our approach this leads to a shift to higher opinion values.

- The present list was generated using review data from Amazon.com. The result is that the words in the list are somehow biased in the sense that the vocabulary is typical for the one used in web platforms. We believe that the list is suitable for applications focused on the analysis of text data retrieved from Web 2.0 platforms but will have problems when being applied to other text data.

4.5.2 Future Work

It is planned to make the list of opinion values for opinion bearing phrases available to the scientific community soon. Some improvements and corrections have to be done before:

- The shift of the opinion values to the positive range (see Section 4.5.1) has to be investigated and corrected.

- Missing phrases can be calculated by using shifter values derived from opinion trees of words with a similar polarity (see Section 4.3.3). The determination of a suitable algorithm was not subject of the present work and, thus, remains to be done.

- The list has to be cleaned manually as it contains some faulty words. The most "prominent" of these words is "amazon", which has been tagged as an adjective after setting it in small letters (see Section 3.2.2).

- Beside adjective-based phrases, also verb- and noun-based phrases are of great interest [37]. Thus, opinion lists for noun and verb based phrases as well as phrases combining several parts of speech will be derived using the algorithm described in this paper.

5. CONCLUSIONS

We introduced an approach of building an opinion value list with both opinion bearing adjectives and adjective-based phrases. Therefore, we used the titles and star ratings of product reviews assuming that both represent a short summary of the writer's opinion. We could show this to be a promising attempt as first results indicate that reasonable opinion values are derived for words and phrases. In principle, this approach allows a fast generation of opinion lists for several languages. Also, domain-specific lexical resources can be generated to a certain extent. The algorithm can easily be extended to other parts of speech so that lists for opinion bearing nouns or verbs can be produced. The list will be made available to the scientific community in the near future.

6. ACKNOWLEDGMENTS

The authors would like to thank all members of the Institute of Information Systems (iisys) of the University of Applied Sciences Hof for many helpful discussions and especially R. Göbel for his great efforts as to the foundation of the institute.
The Institute of Information Systems is supported by the Foundation of Upper Franconia and by the State of Bavaria.

7. ADDITIONAL AUTHORS

Daniel Simon (University of Applied Sciences Hof, email: `ds@simon-ibv.de`).

8. REFERENCES

[1] H. Amiri and T.-S. Chua. Mining Slang and Urban Opinion Words and Phrases from cQA Services: An Optimization Approach. In *Proceedings of the 5th ACM international conference on Web search and data mining*, pages 193–202, 2012.

[2] S. Baccianella, A. Esuli, and F. Sebastiani. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pages 2200–2204, 2010.

[3] J. Brooke, M. Tofiloski, and M. Taboada. Cross-Linguistic Sentiment Analysis: From English to Spanish. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, pages 50–54, 2009.

[4] P. Carvalho, L. Sarmento, M. J. Silva, and E. d. Oliveira. Clues for Detecting Irony in User-Generated Contents: Oh...!! It's "so easy" ;-). In *Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement*, pages 53–56, 2009.

[5] Y. Choi and C. Cardie. Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 793–801, 2008.

[6] S. Clematide and M. Klenner. Evaluation and Extension of a Polarity Lexicon for German. In *Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 7–13, 2010.

[7] A. Esuli and F. Sebastiani. Determining Term Subjectivity and Term Orientation for Opinion Mining. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, pages 193–200, 2006.

[8] A. Esuli and F. Sebastiani. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pages 417–422, 2006.

[9] N. Hu, P. A. Pavlou, and J. Zhang. Why do Online Product Reviews have a J-shaped Distribution? Overcoming Biases in Online Word-of-Mouth Communication. *Marketing Science*, 198:7, 2007.

[10] N. Hu, J. Zhang, and P. A. Pavlou. Overcoming the J-shaped Distribution of Product Reviews. *Communications of the ACM*, 52:144–147, 2009.

[11] N. Jindal and B. Liu. Review Spam Detection. In *Proceedings of the 16th International World Wide Web Conference*, pages 1189–1190, 2007.

[12] N. Jindal and B. Liu. Opinion Spam and Analysis. In *Proceedings of the 1st ACM International Conference on Web Search and Data Mining*, pages 219–230, 2008.

[13] N. Jindal, B. Liu, and E.-P. Lim. Finding Unusual Review Patterns Using Unexpected Rules. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pages 1549–1552, 2010.

[14] M. Klenner, S. Petrakis, and A. Fahrni. Robust Compositional Polarity Classification. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, pages 180–184, 2009.

[15] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. Detecting Product Review Spammers using Rating Behaviors. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pages 939–948, 2010.

[16] B. Liu, M. Hu, and J. Cheng. Opinion Observer: Analyzing and Comparing Opinions on the Web. In *Proceedings of the 14th International World Wide Web Conference*, pages 342–351, 2005.

[17] B. Liu and L. Zhang. A Survey of Opinion Mining and Sentiment Analysis. In C. C. Aggarwal and C. Zhai, editors, *Mining Text Data*, pages 415–463. Springer US, 2012.

[18] J. Liu and S. Seneff. Review Sentiment Scoring via a Parse-and-Paraphrase Paradigm. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 161–169, 2009.

[19] M. Marcus, G. Kim, A. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger. The Penn Treebank: Annotating Predicate Argument Structure. In *Proceedings of the ARPA Human Language Technology Workshop*, pages 114–119, 1994.

[20] G. A. Miller. WordNet: A Lexical Database for English. *Communications of the ACM*, 38:39–41, 1995.

[21] K. Moilanen and S. Pulman. Sentiment Composition. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, 2007.

[22] S. M. Mudambi and D. Schuff. What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly*, 34:185–200, 2010.

[23] A. Mukherjee, B. Liu, and N. Glance. Spotting Fake Reviewer Groups in Consumer Reviews. In *Proceedings of the 21st International World Wide Web Conference*, pages 191–200, 2012.

[24] J.-C. Na, T. T. Thet, C. S. G. Khoo, and W. Y. M. Kyaing. Visual Sentiment Summarization of Movie Reviews. In *Proceedings of the 13th International Conference on Asia-pacific Digital Libraries*, pages 277–287, 2011.

[25] D. Oelke, M. Hao, C. Rohrdantz, D. A. Keim, U. Dayal, L.-E. Haug, and H. Janetzko. Visual Opinion Analysis of Customer Feedback Data. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 187–194, 2009.

[26] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: Sentiment Classification using Machine Learning Techniques. In *Proceedings of the Conference on Empitical Methods in Natural Language Processing*, pages 79–86, 2002.

[27] A.-M. Popescu and O. Etzioni. Extracting Product Features and Opinions from Reviews. In *Proceedings of the Human Language Technology Conference*, pages 339–346, 2005.

[28] R. Remus, U. Quasthoff, and G. Heyer. SentiWS – a Publicly Available German-language Resource for Sentiment Analysis. In *Proceedings of the 7th International Conference on Language Ressources and Evaluation*, pages 1168–1171, 2010.

[29] E. Riloff and J. Wiebe. Learning Extraction Patterns for Subjective Expressions. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 105–112, 2003.

[30] H. Takamura, T. Inui, and M. Okumura. Extracting Semantic Orientations of Words using Spin Model. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 133–140, 2005.

[31] H. Tang, S. Tan, and X. Cheng. A survey on sentiment detection of reviews. *Expert Systems with Applications*, 36:10760–10773, 2009.

[32] P. D. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424, 2002.

[33] L. Velikovich, S. Blair-Goldensohn, K. Hannan, and R. McDonald. The viability of web-derived polarity lexicons. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 777–785, 2010.

[34] U. Waltinger. GermanPolarityClues: A Lexical Resource for German Sentiment Analysis. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pages 1638–1642, 2010.

[35] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of the Human Language Technology Conference*, pages 347–354, 2005.

[36] K. Zhang, Y. Cheng, Y. Xie, D. Honbo, A. Agrawal, D. Palsetia, K. Lee, W.-k. Liao, and A. Choudhary. SES: Sentiment Elicitation System for Social Media Data. In *Proceedings of the 11th IEEE International Conference on Data Mining*, pages 129–136, 2011.

[37] L. Zhang and B. Liu. Identifying Noun Product Features that Imply Opinions. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, volume 2, pages 575–580, 2011.