OPINION MINING/SENTIMENT ANALYSIS

by

ZAKARI ABDUL SHAKUR

REG NO:40238

A dissertation submitted in partial fulfillment of the requirements for the award of

MASTER OF SCIENCE IN COMPUTER SCIENCE

April 2013.

Department of Computing
African University of Science and Technology,
Abuja,Nigeria.

Supervised by:Prof. Amos David
Lorraine University,France

I declare that this dissertation is my own work and that the work of others is acknowledged and indicated by explicit references.

Zakari Abdul Shakur
April 2013.

Opinion Mining /Sentiment Analysis

Dedication

I dedicate this work to the ALMIGHTY ALLAH,

Creator of the entire Universe for granting me the strength,

Wisdom and grace required to complete this work,And to my late

grand mother Hajiya Hamamata J.Abdullahi May her soul rest in peace .

THANK YOU ALLAH!

Acknowledgements

I would like to express my deepest appreciation to my supervisor, Prof.Amos David from the university of Loraine,France. Whom despite his tight schedule squeeze his time to guide and review this work.I also wish to thank the Head of the computer science stream Prof. M.K.Traore who got me started on this exciting journey into the world of computing, especially programing and also for his scholarly assistance. A thank you to Prof. Winston Wole Soboyejo for his sound advice,and everything he did in enhancing my graduate school experience and for leaving an indelible mark in my life in ways that words alone can't fully express. In addition,a thank you to Dr.Mohammed Kabir for his expertise guidance and provision of some literatures to read on this work. Thank you to the committee that sat down to review this work and to all my Professors who taught me during my 18 month sojourn here at the African University of Science and Technology especially Prof. Kola Babalola and Prof.Lehel Csato. You all contributed to the success of this work for which I remain heavily indebted to you.

To the sponsors of my entire Graduate School program, Petroleum Technology Development Fund.(PTDF),I say Thank you, for not only paying the full Tuition for my studies here at AUST but also for granting me an opportunity of a life time to mingle and acquire knowledge from Erudite Scholars.

Special Thanks to Dr.Abubakar Siddique of the Ahmadu Bello University,Zaria. Alh.Rabiu Muhammad(MD Dangote Cement Plc),Alh.Ibrahim C.k,Alh.Adamu Dalhatu(GM.Dangote Cement Plc),and all staff of fleet four,Dangote Cement Plc for being so supportive and understanding.

A big thank you to my lovely parents Hajiya Maimuna Zakari and Alh.Hamid Zakari for the sound upbringing and to my lovely sisters Shafa and Ni'ima.You have been a source of inspiration to me.Also I want to thank my Lovely heart, Maryam Lawal for her Prayers,Love,Patience and understanding who was always there to cheer and lighten me up when the work was getting tougher.A thank you to my room mate(Hafeez Yusuf Hafeez) for his brotherly caring and understanding.

Special thanks should be given to my student colleagues especially those of the CSStream;Yoro,Usman(Walli),Ignace,Doyin,Thabot,Bright,Aligombe,Sahabi,Emmanuel ,Hajara,Nila,Farouk,Patrick,Abderamane,Dagan,Kabir You all were very inspiring and made my stay at AUST a memorable one. To Alh.Aliko Dangote thank you for your heavy financial and moral support.

Finally to everyone who has contributed in one way or the other to the success of this work, I say thank you. Space limitation would not allow me to mention everyone's name here but you know yourselves. May the Almighty God richly bless you in all your endeavours.

Opinion Mining /Sentiment Analysis

Abstract

There has been tremendous growth in the amount of user generated content on the Internet. This is due to rise in social networks and the embrace of web 2.0 or technologies. The Internet is no longer a place for consumption of information only, but a melting point of users interacting on various interests. This project concentrates on a specific type of content – opinionated content. Several review sites exist where users can comment on their experience about a movie, product or service. The review sites allow users to rank their experience of such products or service.

The literature will be reviewed in detail. Several techniques for automatically analyzing such opinionated data will be explored. The focus will be on semantic orientation of the text. Semantic orientation is a measure of how far the opinion contained in the text differs from the group of other words surrounding the text. And will be implemented and compared with other similar work.

Table of Contents

Objectives

This project will establish what sentiment analysis is and define it by putting it into context. It will also establish why it is difficult and distinguish between sentiment classification and subjectivity. Disambiguate fact from opinion and sentiment. Also an algorithm will be implemented with a search engine that has not been implemented before.

Definition of terms

API: Plication programming interface

Corpus: the collection of reviews fact: a thing that is known or proved to be true.

Opinion: view of judgement formed about something

Sentiment: a view or opinion that is held

n-gram: sequence of items / words where n is the limit

Lexicon: lexicon is the list of words or vocabulary of language

supervised learning: approach that requires learning or training to infer knowledge

unsupervised learning: try to find pattern from unlabeled data. Doesn't require training

polarity: could be positive or negative i.e. recommended or not

semantic orientation: evaluative character of a text or group of texts

List of tables

List of figures

1. Introduction: Overview

Sentiment analysis is a technique that allows computers analyse texts from comments, blogs, review aggregation websites and various types of social media to determine opinions about products and services or a domain such as movie reviews. The aim is to extract opinions, emotions and sentiments in the text. The sentiment of the text is then categorized into positive or negative, recommended (thumbs up) or not recommended (thumbs down) or scaled into 1 to 5 star categories. Amazon1 for instance, use star ratings.

Sentiment analysis application areas are for example, a brand tracking what bloggers are saying about a new product or service. As more consumers make purchases online, this potential customers go through reviews left by other customers who have purchased a similar product often basing their decision to buy on the ratings. A five star product is more likely to be preferred over a 2 star product. Tracking this opinions is in the interest of the product manufacturer as well.

However, the application of sentiment analysis is not restricted only to consumer goods sector. Sentiment analysis tasks could be applied to get political opinions of anybody. It could be applied to newspapers stories or editorials. This when applied to different news sources could help highlight different opinion holders in media. This knowledge can then be used for targeted adverts. Also, businesses can track new product perception, detect flames and general brand perception.

Www.amazon.com/uk

2. Sentiment Analysis Literature Review

The web is growing rapidly. The rise of Web applications that promote interactions between users is at the forefront of this growth. Web 2.0 platforms, those sites that fuel this growth include facebook, twitter, review aggregation sites such as epinions and rotten tomatoes have revolutionized the way and amount of data that is created and consumed on a daily basis. The mainly static pages offering little interaction between the users has been transformed to user- driven community based social networks. This also extends to e-commerce sites where users can post comments and opinions about products or services. Forums, wikis and blogs relating to different domains and interest groups generate huge amounts of data as well. This data contains valuable information that could be analyzed for the sentiment and opinions expressed. Both businesses and customers want to get opinions but for different reasons. Businesses can use this information to learn what is liked or not about products. The general sentiment expressed on a recently released product could be used for product or service upgrades or to counter the opinions raised as the case may be. Customers or consumers on the other hand, want to learn of past experiences of others before making decisions to buy or not. People generally seek for opinions of friends and family when thinking of purchasing a good of service. In the days gone by, word of mouth was the only option or adverts on bulletin boards. However, nowadays the e-commerce sites that sell the product also carry reviews by past users of the product. Many other special interest
offer profess recommendations such as www.cnet.com, www.zdnet.com for technology related issues, www.rottentomatoes.com, www.imdb.com for movie and series ratings and other multi- genre sites such as www.epinions.com and www.metacritic.com.

These sites have wide ranging review rating mechanisms. For instance, Amazon lets users leave comments as well as a rating of between 1 to 5 stars. Metacritic.com has colour codes of green, yellow and red as a summary for a review. This serves as a quick indication of the sentiment of the reviewer without actually reading the review. Rottentomatoes.com provides in a nutshell whether a review is fresh (recommended) or rotten (not recommended).

The value attached to such reviews by publishers and producers has led to several criticisms of these sites especially when a review has a good score from reviewer A but receives a lower rating when converted to or “normalized” for another site.

Opinion Mining /Sentiment Analysis

A study conducted by comScore and Kelsey Group (2007) revealed the impact online consumer generated review on the price consumers were willing to pay for a service that is delivered offline. Over 2000 US residents were surveyed in the study. It was revealed that consumers were willing to pay at least 20 percent more for services that receive a 5-star rating (excellent) rather than for the same service receiving a 4-star rating (good). Of those that consulted an online survey, 41% of restaurant reviewers subsequently visited a restaurant. This is closely followed by 40% for hotel reviewers that subsequently stayed at a hotel. The study also found out that reviews generated by users were more influential than professionally sponsored reviews.
This study was conducted in 2007. With the rapid growth of technology and ubiquity of internet connected mobile devices now, it is likely that the number of internet users relying on such reviews has grown significantly.

Independent review aggregation sites also exist. These sites such as UK based which offer advice on a wide range of products and services and conducts extensive tests by independent professionals providing objective information to consumers.

Although most of the sites mentioned above have some form of measure of strength of opinion, much of the information contained in the review has no order to it. The contents of the review is largely unstructured.

A reviewer might compare a product to a former product he owns and offer a comparison of products. This makes it more difficult for a computer to distinguish what is said about product 'a' and what is said about the other product. A human on the other hand, will have no problem in recognizing and associating the sentiment of the opinion holder. This is one of the reasons why sentiment analysis is a challenging task for the computer.
In this work, I hope to apply the PMI-IR algorithm to a group of reviews from different domains and evaluate the performance of the algorithm over those domains. Much of sentiment analysis work has been done on the movie review domain. Tong (2001) developed a system for generating sentiment timelines – this system tracks online discussions about the movies. Turney (2002) and

http://www.which.com

Pang et al. (2002) all involved movie reviews although the work by (Turney, 2002) also highlighted why classifying movie reviews tend to be more difficult than other domains for example automobiles, travel reviews etc. This is because many techniques of sentiment analysis rely on the words (lexicon) contained in the review to determine which statement indicate positive sentiment and negative sentiment. The problem with this is that a lot of the reviews contain sarcasm as well e.g "That movie was an excellent waste of time". This will be explored in a subsequent section

2.1 Structured and Unstructured data

Although the data contained in the reviews is highly valuable, mining this data for opinions is a very challenging task due to the fact that it is largely unstructured. Structured information is typically found in databases records where the semantics of the data is known. Sales figures per quarter by region can be calculated easily by querying such databases. This data often have a structure and format to it.
Unstructured data on the other hand, refers to the information that is generally out of the organisations control. This include text of comments, blog posts, news articles etc. Difficulties arise, however, when mining this huge volume of disorganized data. Therefore, sentiment analysis is an attempt to find value and meaning from this huge amount of data.

2.2 A note on Terminology (fact, opinon, sentiment, sentiment analysis, opinion mining)

According to the Oxford online dictionary3, a fact is " a thing that is known or proved to be true. Opinion is a " view of judgement formed about something" while sentiment is " a view or opinion that is held".

Sentiment analysis is a wide and multidisciplinary field. There are many definitions and titles that all broadly refer to the same topic, for example, opinion mining. Sentiment or opinion itself varies and the level or depth of sentiment at which the analysis or mining occurs varies.

In literature, sentiment analysis goes by different names – opinion mining for instance in Morinaga

3 http://oxforddictionaries.com

et al (2002), Dave et al (2003) and Grefenstette et al (2004). Sentiment analysis in Nasukawa and Yi (2003) and Pang and Lee (2004). Sentiment extraction in Pang et al (2002) and Fei et al (2004).

2.3 Fundamentals of Sentimental Analysis and key concepts

In this section, the different approaches to solving major tasks of sentiment analysis and classification problems will be discussed. The main task or technology in sentiment analysis is classification. This entails classifying text as either positive or negative. The task of subjectivity classification will also be discussed. There are also approaches utilising machine learning algorithms with a good degree of accuracy and unsupervised methods as well.

2.3.1 Semantic Orientation

It is necessary to introduce semantic orientation before highlighting its various applications in literature. Turney (2001) defines semantic orientation as the evaluative character of a word. The semantic orientation has direction as well as intensity. Words that kindle desirable states have positive orientation eg excellent, praise and negative show undesirabilty – poor, bad etc. The semantic orientation or polarity of a word indicates the direction the word deviates from the norm for its semantic group (Lehrer, 1974).

2.3.2 Sentiment Classification

Sentiment classification essentially takes text containing opinion as input and processes it into two classes positive and negative. A positive opinion shows that the text is desirable by the reviewer and negative not so. Much work on sentiment classification applies to the travel or movie reviews domain where the system produces positive and negative reviews for a travel destination or movie.
This is possible by extracting sentiment words that are associated with positive or negative opinions in the natural language. Words such as excellent, nice, good and nice etc. are associated with

positive sentiments while bad, poor, negative, horrible etc. indicate negative opinions.

However, other reviews, commentary or news do not necessarily contain negative or positive sentiment. It could be reporting news or statement of fact. For example, 'there was a fuel price hike' is a statement of fact. News stories about stock price changes of a company have also been classified by (Koppel and Shtrimberg 2004) into good or bad news.

The task of sentiment classification occurs at different levels of a document or text. The document- level evaluation considers the whole review as an entity and does not go into detail of what are the specific and favourable features of the reviewed item.
The sentence-level breaks down a document into sentences and classifies the sentences individually as either expressing negative or positive opinion.

2.4 Document-level approach to Sentiment analysis: Unsupervised Approach

Document level opinion analysis classifies the whole document's overall sentiment as either negative or positive. Both machine learning and unsupervised approaches will be examined in this section.

The unsupervised approaches to sentiment analysis do not rely on labelled data or training of algorithm. This approach determines subjectivity or opinion of the text based on words that indicate sentiment like adjectives (beautiful, excellent etc.) as they convey a high degree of opinion.

Turney (2002) presented an unsupervised algorithm for classifying reviews as recommended or not. Past work by Hatzivassiloglou and McKeown (1997) has shown that adjectives are good predictors of subjectivity. Although the adjectives may indicate subjectivity, it is not sufficient in some contexts to determine semantic orientation. This is illustrated with the adjective “unpredicatble” which would have a negative orientation in a car review when it occurs in a phrase “unpredictable steering” but in a movie review an “unpredictable plot” could be delightful for the audiences. To solve this problem, a second word is extracted along with the adjective to provide context. This

Opinion Mining /Sentiment Analysis

The algorithm consist of three steps:

Step 1: The first step of the algorithm takes a text document as input. This document is the review. The document is processed by a part of speech tagging software (Brill, 1994) to extract words that conform to the patterns in Table 1. The JJ tags indicate adjectives, the NN and NNS for noun(s), the RB for adverbs, and the VB tags are verbs.

Table 1. Pattern of tags for extracting two-word phrases from reviews (Turney 2002)

First Word	Second Word	Third Word (Not Extracted)
1. JJ	NN or NNS	Anything
2. RB, RBR, or RBS	JJ	Not NN nor NNS
3. JJ	JJ	Not NN nor NNS
4. NN or NNS	JJ	Not NN nor NNS
5. RB, RBR, or RBS	VB, VBD, VBN, or VBG	Anything

Example 1: In the sentence “great storyline about being true to yourself”, “great storyline” will be extracted as it satisfies the first pattern. The second pattern pattern means that two consecutive words are extracted if the first word is an adverb and the second word is an adjective but the third word can not be a noun even though the third word is not extracted.

Step 2: In the second step, the semantic orientation of the phrases is calculated. The pointwise mutual information is a measure of the strength of the association between words (Church and Hanks, 1989). The semantic orientation of the phrase is calculated based on the pointwise mutual information algorith. It is defined as follows (Church and Hanks, 1989):

$$\text{PMI}(\text{word1}, \text{word2}) = \log(2) \frac{p(\text{word1}) \wedge p(\text{word2})}{p(\text{word1})\, p(\text{word2})} \qquad (1)$$

From equation 1 above, “p(word1 and word2) is the probability of word1 and word2 co-occur. If the words are statistically independent, then the probability that they co-occur is given by the product p(word1)p(word2). The ratio between p(word1 & word2)

$$PMI(word1, word2) = \log_2 \frac{p(word1 \wedge word2)}{p(word1)\, p(word2)} \qquad (1)$$

and p(word1) p(word2) is thus a measure of the degree of statistical dependence between the words. The log of this ratio is the amount of information that we acquire about the presence of one of the words when we observe the other".

The Semantic Orientation (SO) of a phrase, is calculated based on its association with the positive reference word "excellent" and negative reference word "poor" .

$$SO(phrase) = PMI(phrase, excellent) - PMI(phrase, poor). \qquad (2)$$

The semantic orientation value will be positive if the phrase is more strongly associated with "excellent" and negative if the phrase is more associated with "poor". Therefore, according to the formula words that co-occur more frequently with "excellent" than "poor" will have a positive semantic orientation and will probably be positive.

The PMI is calculated by issuing search queries to a search engine and collecting hit counts. This is why the algorithm is PMI-IR (Information retrieval) . This is the information retrieval part of the algorithm. AltaVista search engine was used because of its near operator which constrains the search to pages that contain the phrase and the reference word within a distance of ten words, either to the right or left.

Let hits(query) be the number of hits returned. Equation 2 can be rewritten as:

$$SO(phrase) = \log_2 \frac{hits(phrase\ NEAR\ excellent)\, hits(poor)}{hits(phrase\ NEAR\ poor)\, hits(excellent)} \qquad (3)$$

Step 3: Given a review, the algorithm computes the average SO of all the relevant phrases in the review. The review is classified as as recommended if the average SO is positive and otherwise if the SO is negative.

The PMI-IR algorithm relied on Alta-Vista's NEAR operator for queries which has since been deprecated. The PMI-IR was applied to various domains and the accuracy ranges by domain as well. The author reported that the automobile domain got the highest accuracy with 84% then banks with 80% and movies the least with 65% accuracy.

Turney and Littman (2003) slightly modified the algorithm to include more seed words. This approach they called semantic orientation by association SO-A. The new words were selected for their insensitivity to context. The positive words include {good, nice, excellent, positive, fortunate, correct and superior} while the negative words chosen are {bad, nasty, poor, negative, unfortunate, wrong and inferior}.

The work most closely associated with Turney (2002) is Tong's (2001) system for tracking the sentiments expressed on movies in forums over time. The system built a specific lexicon for the movie domain. These phrases were manually tagged into two groups that indicate negative or positive sentiment. The phrases “great acting”, “wonderful visuals”, “terrible score” and “uneven acting” is clearly movie review domain. Such domain specific phrases will not do well in automobile reviews for instance. The disadvantage of this approach is that a new lexicon has to be built if it to be applied to another domain.

2.5 Document-level approach to Sentiment analysis: Machine Learning Models

In contrast to Turney et al's unsupervised approach, Pang et al (2002) applied machine learning techniques to sentiment classification problem. As the results from Turney (2002) showed that movie reviews was the most difficult for sentiment classification as compared to automobile, banks and travel destination domains, they also use movie review data to show that machine learning techniques perform better than human-produced baselines. In order to produce this baselines, two graduate students read reviews and identified those words that are indicators for negative and positive sentiment and rely on the words alone for classification. This is in line with the fact that certain words such as adjectives show sentiment.

The tables below highlight the proposed word list.

	Proposed word list	*Accuracy*	*Ties*
Human 1 .	Positive: dazzling, brilliant, phenomenal, excellent, fantastic negative: suck, terrible, awful, unwatchable, hideous	58%	75%
Human 2	Positive: gripping, mesmerizing, riveting, spectacular, cool, awesome, thrilling, badass, excellent, moving, exciting negative: bad, cliched, sucks, boring, stupid, slow	64%	39%
Human 3 + stats	Positive: love, wonderful, best, great, superb, still, beautiful negative: bad, worst, stupid, waste, boring, ?, !	69%	16%

Table 2: Human-produced baselines Pang et al (2002)

This was applied to 700 negative and 700 positive known movie reviews. The decision procedure counts the number of seed words in a given document to determine sentiment. The results from human 1 and two respectively is 58% and 64%. The third word list was created after further examination of the corpus. This also includes punctuation marks “?” and “!”. This list performed better than the first two proposed.

Interestingly, the Human 3 seed words have no tie with the Turney's “excellent” or “poor” seed words and reports better accuracy of 69% as against PMI-IR algorithms 65.83% in the movie review domain.
In order to improve the third baseline of 69%, machine learning methods were applied. Specifically, Naive Bayes(NB) classifier, maximum entropy(ME) classification and support vector machines (SVM) classification was tested to see which will produce the best result. The table below provides a summary of the results

	Features	*No of features*	*Frequency or presence?*	*NB*	*ME*	*SVM*
1	Unigram	16165	Freq	78.7	N/A	72.85
2	Unigram	16165	Pres.	81.0	80.4	82.9
3	Unigram+ Bigram	32330	Pres.	80.6	80.8	82.7
4	Bigrams	16165	Pres.	77.3	77.4	77.1
5	Unigram+ POS	16695	Pres.	81.5	80.4	81.9
6	Adjectives	2633	Pres.	77.0	77.7	75.1
7	Top 2633 Unigram	2633	Pres.	80.3	81.0	81.4
8	Unigram+ Position	22430	Pres.	81.0	80.1	81.6

Table 3: Results Pang et. al. 2002

From the table above, 1 and 2 with unigram features have least accuracy of 72.85%. This shows that a supervised approach with only one word or No 7 with adjectives achieving 75.1% clearly performs better than the unsupervised approach of PMI-IR algorithm where two words were chosen to provide context.

Interestingly, Dave et. al. (2003) found out that bigrams and trigrams can outperform unigrams in some cases.

2.6 Other Related Document Level Approaches

Huettner et. al. (2000) proposed a general purpose approach to document analysis and management. They referred to the approach as “fuzzy typing for document management”. This method extracts words from a document and compute its affect emotion. An effect lexicon was built for about 4000 words. Another set of seed words were then created. These seeds words include “anger, happinness, fear, superiority, violence, death”. The rest of the words were then assigned to the categories they most closely fit. The associations are then graded for the degree of relatedness and the intensity of the association. For example:

"gleeful" adj happiness 0.7 0.6

"gleeful" adj excitement 0.3 0.6,

gleeful here is an adjectives that shows some level of happiness. The score of 0.7 is the measure of centrality on the happiness scale. There is also a relationship with excitement but not as strong as happiness. Gleeful has a measure of 0.3 on the excitement scale. The intensity in both cases is 0.6. These measures are on a scale of 0 – 1. The centrality and intensity is not universal to all domains. Such approach requires training. Therefore this is not an unsupervised algorithm. The centrality and intensity must be determined for the domain as it is subjective. A hand lexicon must be built for movie review for instance. The positive (excellent, superior, good, etc) and negative (nasty, wrong, poor) suggested by Turney et al. (2002) semantic orientation by association will be appropriate.
This words were chosen for their neutrality. However, this has not been evaluated yet.

2.7 Sentence - Level

Sentence level approach focuses on the sentences individually. The are two approaches to sentence level analysis. Sentence level subjectivity classification determines whether a sentence expresses an opinion or statement of fact. Subjective statements are opinions of the holder and the opinion expressed may be positive or negative. Objective statements are merely statements of fact. Sentence level classification also rely on words and phrases as does the document level approach of Turney (2002).

2.7.1 Subjectivity Classification

Wiebe et al. (2000) investigated the task of distinguishing factual information from sentences that contained opinions and personal commentary. This task is applicable to news reports, political opinion holders and internet forum members. Subjective sentences convey the authors evaluations, opinions, emotions as well as speculations. This can be compared with a second author who reports only factual information and the interpretations of the information is left to the readers. For example,

1. “The sound quality of the phone is good enough for me” - Subjective sentence.
2. “The has an 8.0 mega pixel camera with optical zoom and flash” - Objective sentence.

..

Many reviews contain factual and subjective information that adds to the noise which makes the task of classifying the sentiment of the document.

2.7.2 Sentence-level classification methods

Meana et al.(2007) work on sentence-level classification concentrated on the effect of conjunctions and sentence constructions. A conjunction is a part of speech that connects words, phrases and sentences together. They are important because they can change the orientation or express the exact opposite of the first idea expressed.

For example - “Mark's wife is a good interior designer, but their house is messy”. Here good interior designer has a positive orientation but the second part of the sentence linked by “but” is a negative.

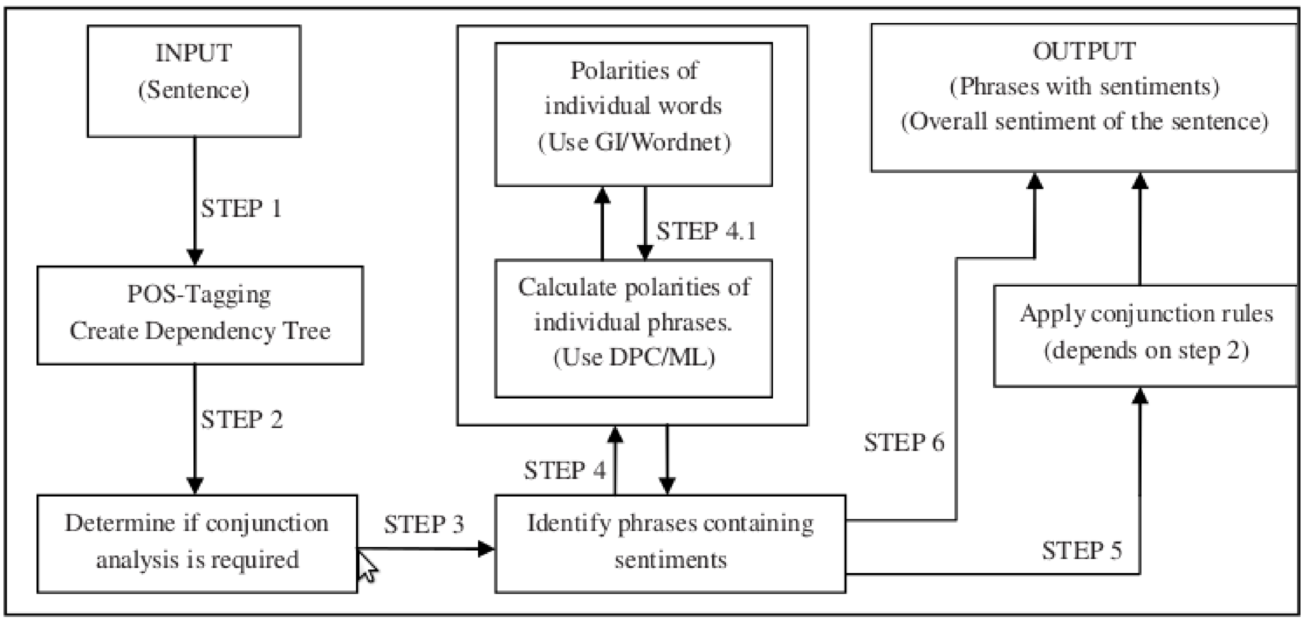


Figure 1: Sentiment classification steps (Meana et. al. 2007)

The process takes a sentence as input. The sentence is then passed to Stanford Lex-Parser 4 which tags the parts of speech and also creates a dependency tree. The phrases containing sentiment are identified if a conjunction is identified. However, the semantic orientation of the phrases are calculated by using the General Inquirer word list. If the word is not on the list then Wordnet is used to calculate the semantic orientation of the phrase.

4 http://nlp.stanford.edu/software/lex- parser.shtml

The General Inquirer (GI) (Stone et. al. 1966) is a tool that maps a text file counting words that correspond to a purpose-built dictionary. These categories are varied and can be built anew. For instance, a “negative” category will contain words that is closely associated with. When a text a mapped, all the instances will be identified and counted. The GI can be applied to sentiment analysis when appropriate categories are built. A positive and negative word list can be built for this purpose. The GI was not originally built for natural language processing or text analysis. Rather, it was applied in social science content analysis research.

The words in each category are giving weights which are then tallied. The result determines the orientation of the document. Meana et. al. (2007) approach will alternatively use Wordnet5(Miller et al.1993; Fellbaum et al. 1993). when there are no hits on the general inquirer list. The dataset used was car reviews achieving accuracy rate of at least 62% was achieved after conjunction analysis for sentences that contained conjunctions.

Kim et. al. (2004) defined opinions as having four factors – topic, holder, claim and sentiment. An opinion has a holder that believes a claim about a topic and may or may not associate a sentiment to the topic. For example, given the topic (e.g.,“Should abortion be banned?”), and the set of texts on the topic, the system will find the sentiments expressed and identifies those who expressed the sentiment. The algorithm first identifies the words (adjective, verb, nouns etc) from sentences that contain both topic phrase and the opinion holder. The polarity of the words containing sentiment are calculated individually and summed up.

The approaches present so far are mostly been applied to product reviews. There are other attempts to distinguish fact from opinions. In the media for instance, news stories generally contain more facts than opinion. This is because it is events that have happened or the stock market indices numbers etc that is reported. Editorials however, contain commentary and perspective given from the point of view of the paper. One such approach was proposed by Yu and Hatzivassiloglou (2003).
One of the motivations for the approach is in information retrieval domain where facts are needed as against subjective information. Another motivation is to solve a question like “What are the reasons for the US – Iraq war?”

The authors hypothesize that opinion sentences are more similar to other opinion sentences than

5 Wordnet.princeton.edu

factual sentences. This “similarity approach” is measured with information retrieval queries and SIMFINDER (Hatzivassiloglou et. al. 2001). The SMFINDER measures similarity based other words, phrases and Wordnet. Words that are similar will have the same orientation. The average of the similarity values provided by SIMFINDER is taken. Finally the sentence is assigned to the category with the highest average. Having identified the words that have sentiment, the semantic orientation of the words is then calculated. The use of known positive and negative seed words is explored further. Turney(2002) proposed “excellent” and “poor”, while the seed words were extended to seven positive and seven negative in (Turney et. al. 2003). Hatzivassiloglou and McKeown (1997) have already built a list of 1336 adjectives with 657 classified as positive and 679 classified as negative. Multiple seed words were chosen to determine if the accuracy rate will be improved. The part-of-speech tags used were also extended from adjectives, adverbs, nouns, verbs and “adjectives and adverbs” to “adjective, adverbs and verbs” and “adjectives, adverbs, nouns and verbs”. At sentence level, the accuracy of over 90% was achieved.

Riloff and Wiebe (2003) contend that in order to present a system that detects subjectivity well, relying on hand built lexicons is not enough. This is due to the fact that many subjective words do not occur frequently in text. Language is also full of figure of speech and idioms that are well understood by the general public. They propose a bootstrap process of a high precision classifier that learns patterns of subjective and objective expression text from unlabelled data. The classifiers also use the adjective lists from (Hatzivassiloglou and McKeown 1997) and other preset lists of part of speech that indicate subjectivity and objectivity as well (Levin, 1993; Ballmer and Brennenstuhl, 1981) etc.

The first classifier “HP-Sub” extracts from the corpus sentences that are subkective while the second classifier “HP-Obj” searches for sentences that are objectives. Only those sentences identified with high confidence are extracted. The authors report that 91% of the sentences selected are subjective.

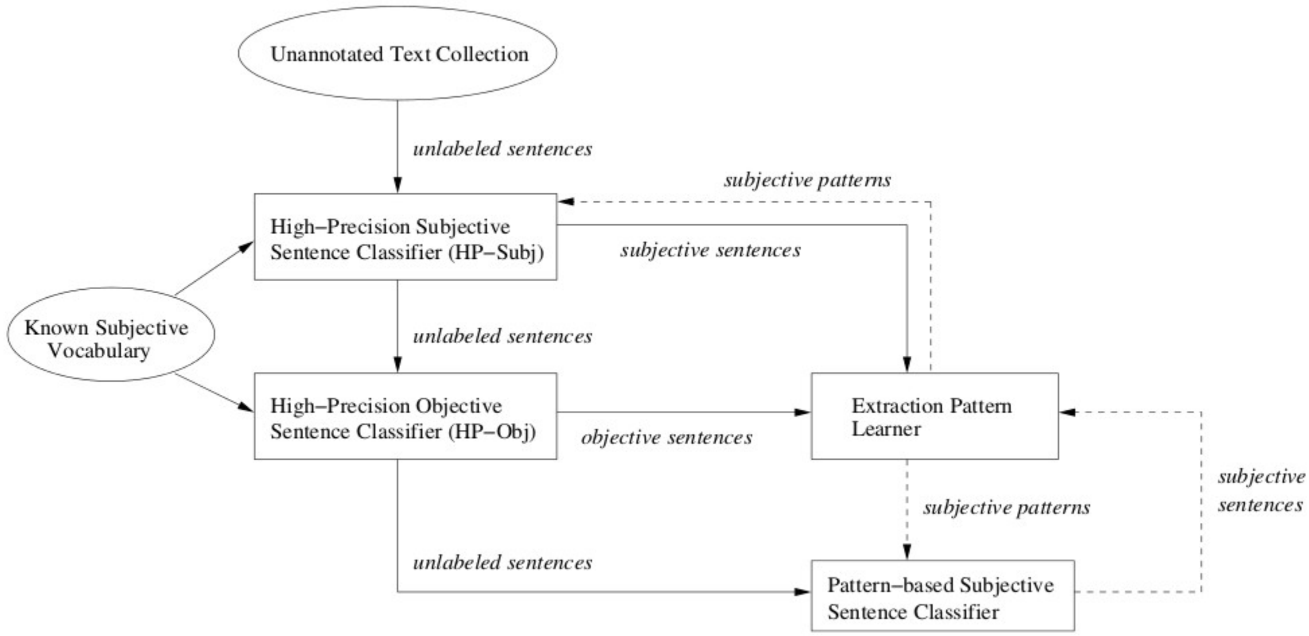


Figure 2: Bootstrapping Process (Rieloff and Wiebe 2003)

As the algorithm is supervised, the second step involves training. The high-precision classifiers generates large sets of subjective and objectives texts automatically. The patterns (figure 3) of the sentences extracted are then used to grow the training set.

SYNTACTIC FORM	EXAMPLE PATTERN
<subj> passive-verb	<subj> was satisfied
<subj> active-verb	<subj> complained
<subj> active-verb dobj	<subj> dealt blow
<subj> verb infinitive	<subj> appear to be
<subj> aux noun	<subj> has position
active-verb <dobj>	endorsed <dobj>
infinitive <dobj>	to condemn <dobj>
verb infinitive <dobj>	get to know <dobj>
noun aux <dobj>	fact is <dobj>
noun prep <np>	opinion on <np>
active-verb prep <np>	agrees with <np>
passive-verb prep <np>	was worried about <np>
infinitive prep <np>	to resort to <np>

Figure 3: Synatictic form and sample pattern extracted (Rieloff and Wiebe 2003)

The results from the experiment carried out on foreign news documents show that the syntactic form of the extraction worked well with a precision range of over 70%.

2.8 Feature-based Opinion Summarization

This approach focuses on specific features of a product to and extracts opinion on the feature sets to determine the sentiment expressed. For example, a review of a mobile phone will that includes features such as large screen size, good build and poor battery life has mixed reviews. Battery life specifically is not reviewed favourably although the overall reviewer may recommend this particular product.

Recent work by (Popescu and Etziomi, 2005) introduced OPINE, an unsupervised information extraction system that identifies product features and opinions on the features by an opinion holder. It also determines the opinion polarity and strength of the opinion.

2.9 Summary

Sentiment analysis will continue to be very valuable as the web grows. More content is created daily within the web and the content is rich with valuable information. We have also seen the power of the spoken language and it's subtleness that makes it difficult for machines to fully understand the human language. The techniques and approaches though challenging are getting better with the day as more powerful machines and techniques are developed.

3 Methodology

This section discusses the experiment and implementation of a review classification system with the aid of web searches. Also, the motivation for choosing the system of implementation will be revealed.

3.1 Data sets

The data set used for any experiment is very important. It has been shown that movie review mining is the most difficult due to noise it contains. This noise is generated by reviewers often talking about the actors' performances as compared in earlier films and many numerous other details other than the actual film itself.
Past research has produced many datasets that are publicly available. The movie review data sets produced by (Pang et al 2002) which is freely available on the internet6 will be the source for movie review data. However, the version 2.0 of the dataset was first used in 2004. The dataset consists of movie reviews from IMDB7 archives. All forms of ratings originally used by the IMDB have been stripped leaving only plain text.
The second dataset containing automobile reviews, phones reviews and cookware is from the work by Taboada and Grieve(2004). The reviews were originally posted to the Epinions website and divided into different categories of 25 positive reviews and 25 negative reviews.The ranking were as originally posted by the reviewers.

3.2.1 Part-of-speech tagging

The task of classifying a document based on pointwise mutual information and information retrieval starts at the tagging phase. The dataset was tagged with the aid of Brill tagger8.

---

6 http://www.cs.cornell.edu/People/pabo/movie-review-data/

7 http://reviews.imdb.com/reviews

8 http://cst.dk/online/pos_tagger/uk/index.html

The automobile reviews, phone reviews and movie reviews were tagged. Approximately 3300 phrases were tagged across all the domains.

Domain	Average Phrases
Automobiles	1607
Phones	952
Movies	733

The second step is to estimate the semantic orientation of the words. For this, various approaches and possibilities were tried and tested .

3.3 API Availability

Turney (2002) implemented with AltaVista search engine. This is because it has a NEAR operator. This has since been deprecated. There is a need to explore other alternatives.

3.3.1 Yahoo BOSS API

Yahoo provides a service that allows developers to integrate yahoo search into their web or desktop applications. The API supports various search operations. The following boolean operations are supported – AND, OR (|, &)9. However, the NEAR operator is not supported. Previous work by (Turney 2001) has also shown that NEAR performs better than the AND operator for proximity queries. Therefore using the API is not suitable for queries.

3.3.2 Google API

Google also does not currently support near proximity search. There was a tool, Google API proximity search (GAPS)10 which was based on the SOAP Search API. The SOAP Search API11 is now deprecated, therefore GAPS does not work. The newer api – Ajax Search API does not support the NEAR operator.

9 http://developer.yahoo.com/search/boss/boss_api_guide/v2_univer_api_args.html

10 http://www.staggernation.com/cgi-bin/gaps.cgi

11 http://googlecode.blogspot.com/2009/08/well-earned-retirement-for-soap-search.html

3.3.3 Google Web1T 5 grams12

Google recently released a data set contain English n-grams and the frequency they occur on the web (Franz and Brants, 2006). The dataset consists of a unigram (a word) to 5-grams (five words). Only those n-grams appearing more than 40 times were included in the dataset. Also, after discarding words appearing less than 200 times, the total unique words in the data set is equal to 13,588,399.
The dataset was released in archive format. The data needs to indexed and query interfaced built. A web interface is available13. The interface is provided by Evert 2005.
Sample data and frequency;

3-gram data
glass bottle markets 2169
glass bottle like 1347

4-gram data
glass bottle filled with – 693
glass bottle or flask – 351

5-gram data
glass bottle saves enough energy 187
glass bottle and plastic container 442

As the data set contains upto 5 n-grams, the queries were adapted to form of ordered searches. For example, the phrase or bigram “good time” could be queried in the following format.

12 http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13

13 http://cogsci.uni-osnabrueck.de/~korpora/ws/cgi-bin/Web1T5/Web1T5_freq.perl

	Pattern	Example	Hits
1	“phrase * * poor”	Good time * * poor	47
2	“phrase * poor**”	Good time * poor *	0
3	“phrase * poor *”	Good time poor * *	0
4	“poor * * phrase”	Poor * * good time	0
5	“ poor * phrase *”	Poor * good time *	0
6	“* * poor phrase”	* * poor good time	0
		Total hits near Poor	47
1	“phrase * * excellent”	Good time * * excellent	54
2	“phrase * excellent *”	Good time * excellent *	0
3	“phrase excellent * *”	Good time * excellent **	0
4	“excellent * *phrase”	excellent * * good time	179
5	“* excellent phrase *”	excellent * good time *	443
6	“** excellent phrase”	* * excellent good time	0
		Total hits near excellent	676

Table 4: WEB1T5 hits

Applying the formula

$$SO(\text{phrase}) = \log(2)\frac{\text{hits}(\text{phrase NEAR excellent})\,\text{hits}(\text{poor})}{\text{hits}(\text{phrase NEAR poor})\,\text{hits}(\text{excellent})} \qquad (4)$$

The semantic orientation of the phrase “good time” is 1.0590707229. A positive number indicates positive orientation. This means that the phrase “good time” is more closely associated with “excellent” than “poor”.

Many other queries were run on the WEB1T5 queries but it turned out to be inappropriate for sentiment analysis. For example, the phrases “drunk director”, “horrible flick”, “unflushed toilet”, “fully entertain”, “fully enjoy” and “crack open” do not occur anywhere near “excellent” or “poor” within the WEB1T5 dataset. The review with this phrases is therefore misclassified.

3.3.4 Yahoo Ordered Search

Proximity search looks for documents where at least two keywords co-occur within a distance in the document. It's been shown that words co-occurring around each other are more likely to be related. Yahoo seach engines appears to support ordered search. For example the query “good * * house” and “good house” has no difference in both hit counts and results. The order is not adhered to by the search engine. Both queries return 1,860,000,000.

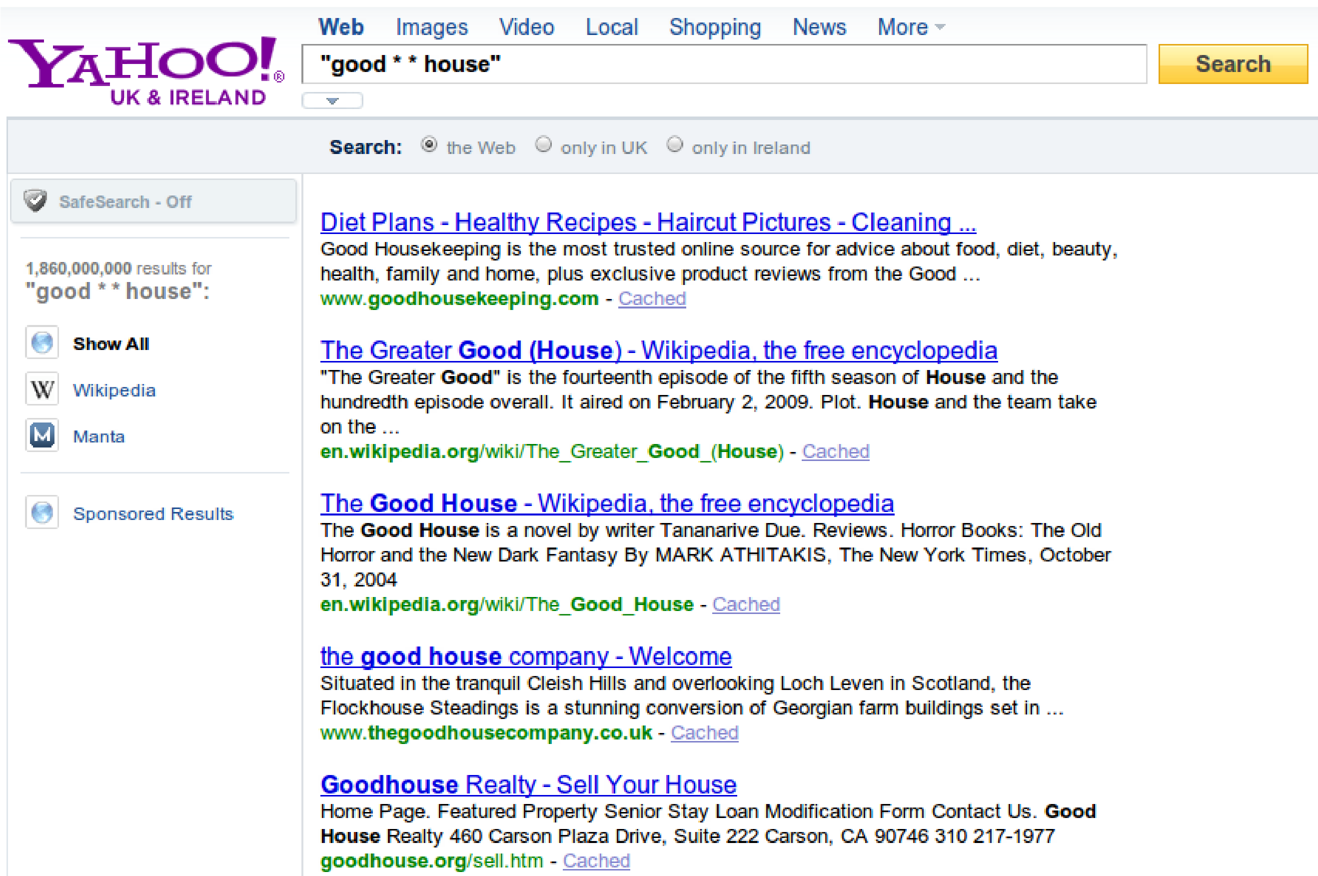


Figure 4: Yahoo search “good * * house”

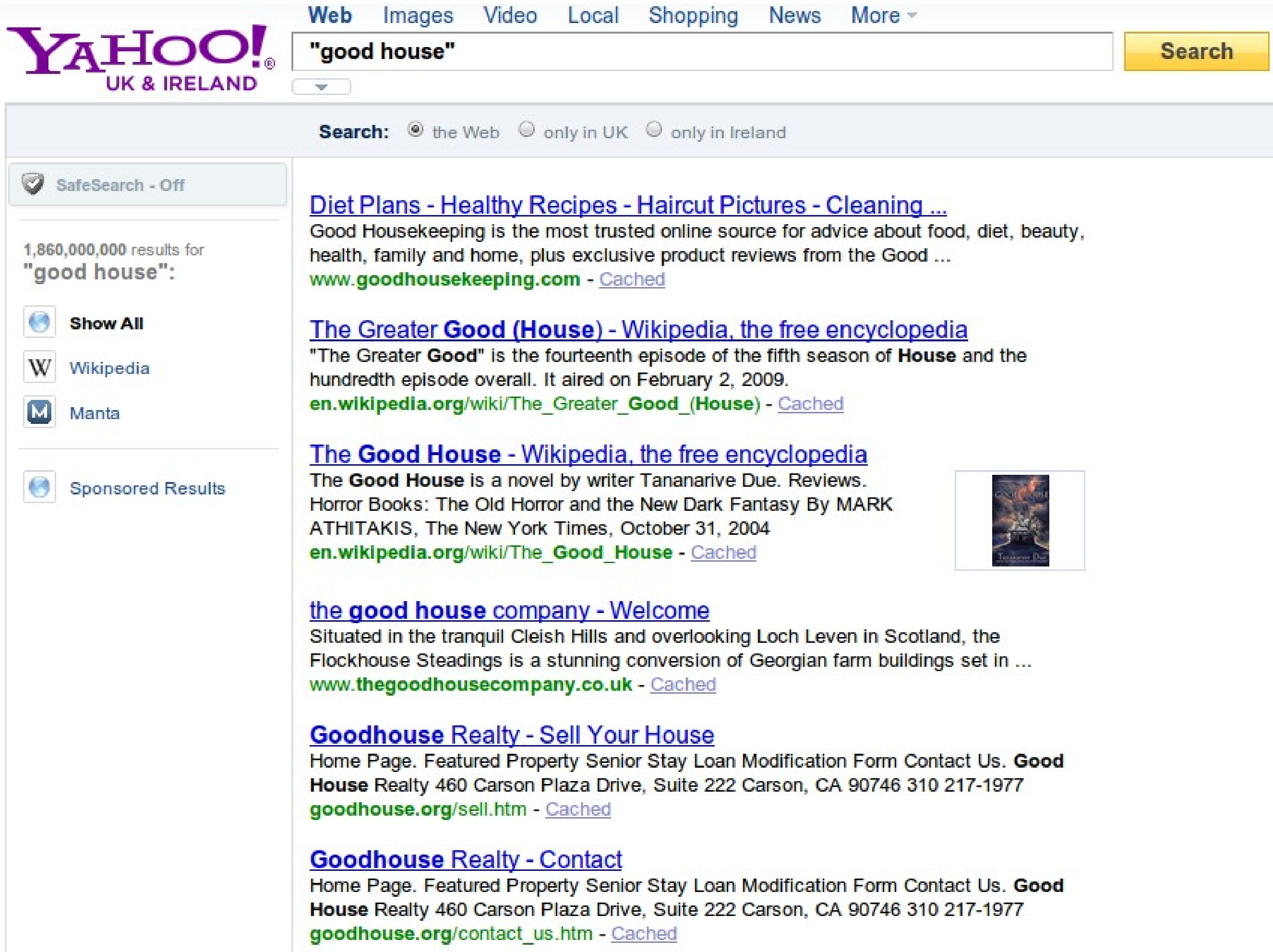


Figure 5: Yahoo query “good house”

3.3.5 Google Ordered Search

The following proximity queries were issued to the Google search engine - “good house * * poor”. Google tries to first match the order and then return the results that 'closely' match the query.

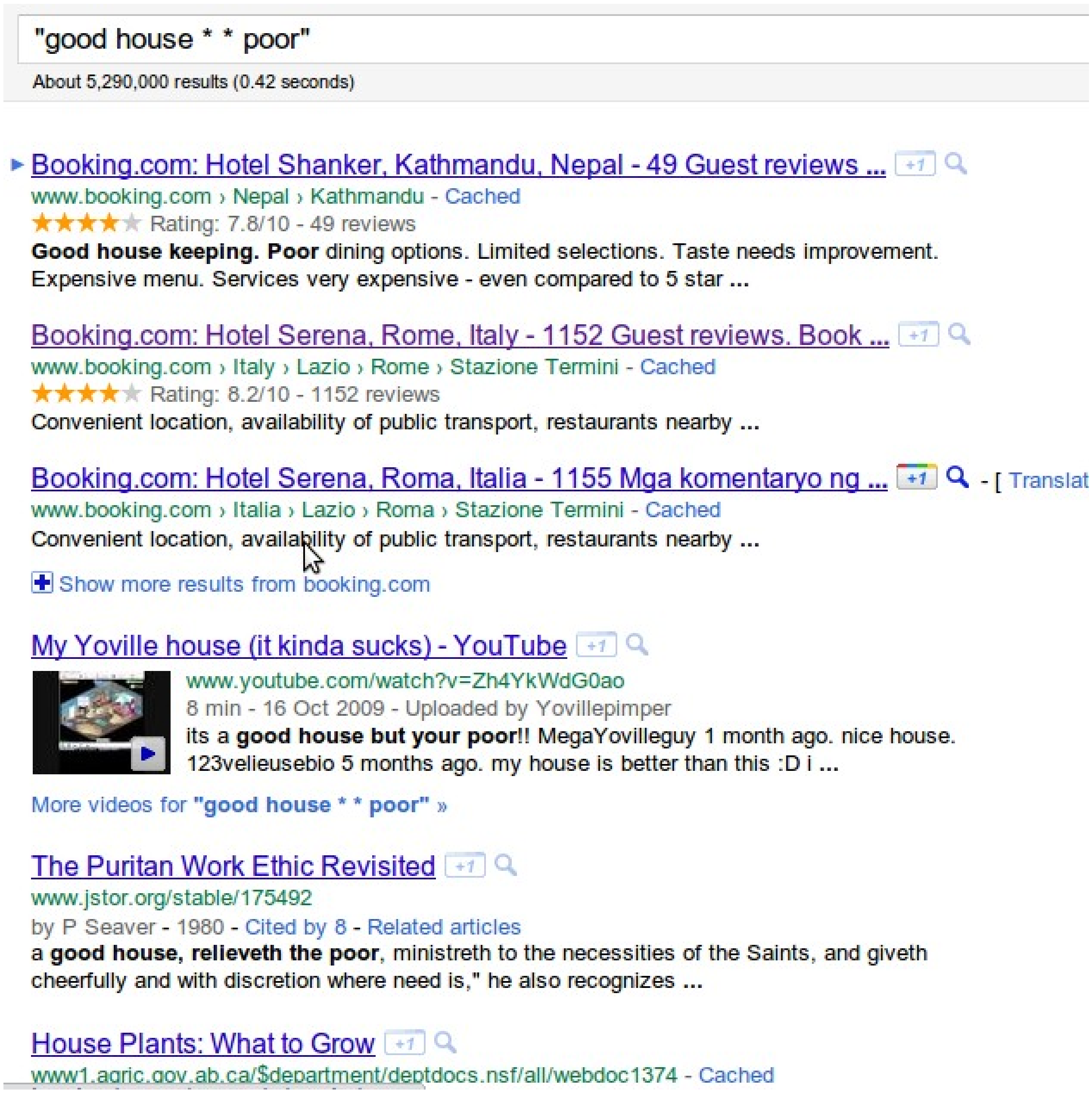
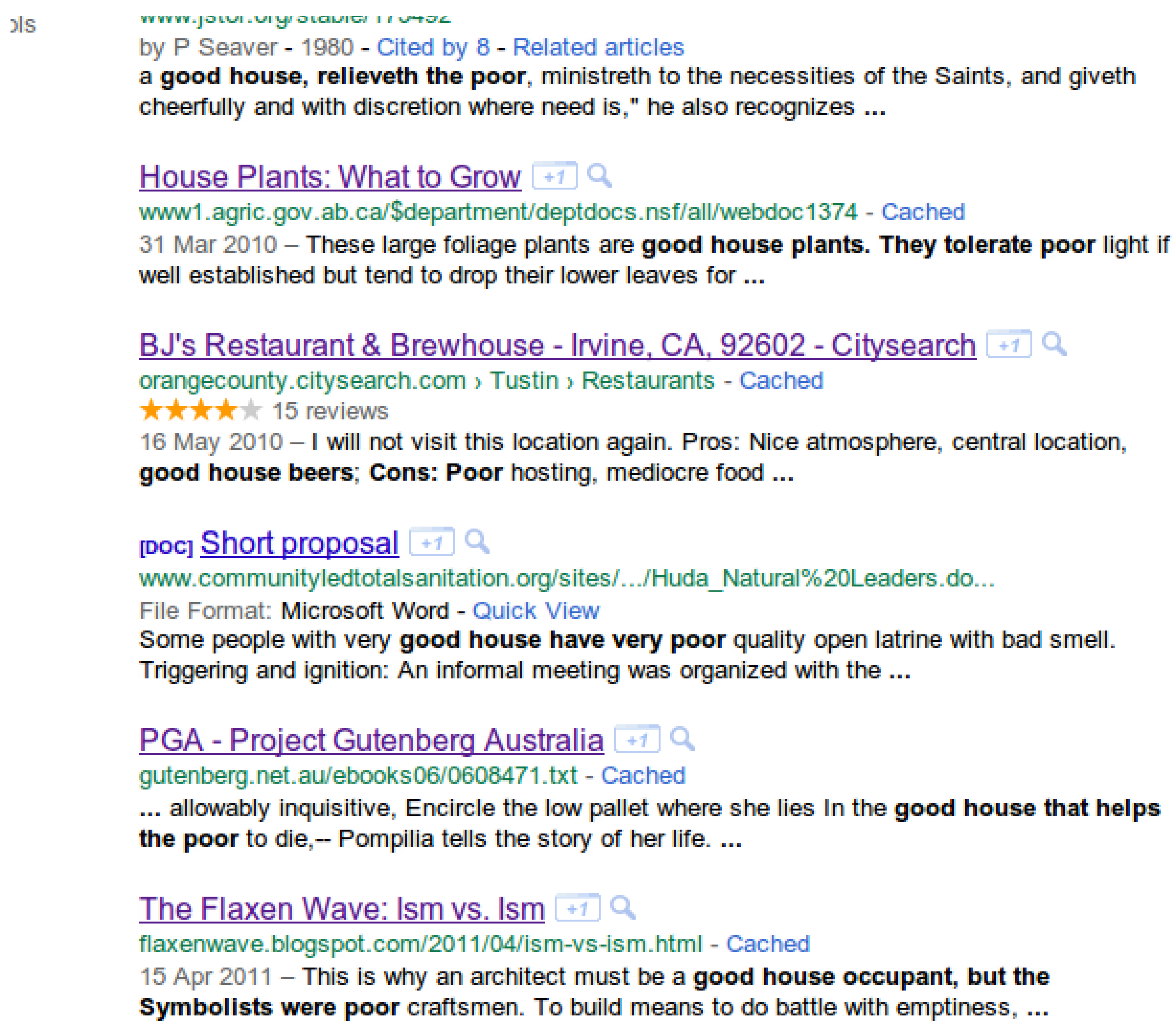


Figure 6: Google query “good house * * poor”

The first result is most closely match, ignoring the “.” in front of keeping - “good house keeping. Poor”. The next two do not match the query at all. It only contains the keywords in the document but not closely related. Also, in the figure below, the proximity constrain is ignored by the search engine. The last result has five words in between - “good house occupant but the symbolists were poor”

From the results of the queries, it is clear that such hit counts can not be relied upon.

Also the lack of support for NEAR operators in both search engines makes them unsuitable for proximity queries



by P Seaver - 1980 - Cited by 8 - Related articles
a **good house, relieveth the poor**, ministreth to the necessities of the Saints, and giveth cheerfully and with discretion where need is," he also recognizes ...

House Plants: What to Grow
www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/webdoc1374 - Cached
31 Mar 2010 – These large foliage plants are **good house plants. They tolerate poor** light if well established but tend to drop their lower leaves for ...

BJ's Restaurant & Brewhouse - Irvine, CA, 92602 - Citysearch
orangecounty.citysearch.com › Tustin › Restaurants - Cached
15 reviews
16 May 2010 – I will not visit this location again. Pros: Nice atmosphere, central location, **good house beers**; **Cons: Poor** hosting, mediocre food ...

[DOC] Short proposal
www.communityledtotalsanitation.org/sites/.../Huda_Natural%20Leaders.do...
File Format: Microsoft Word - Quick View
Some people with very **good house have very poor** quality open latrine with bad smell. Triggering and ignition: An informal meeting was organized with the ...

PGA - Project Gutenberg Australia
gutenberg.net.au/ebooks06/0608471.txt - Cached
... allowably inquisitive, Encircle the low pallet where she lies In the **good house that helps the poor** to die,-- Pompilia tells the story of her life. ...

The Flaxen Wave: Ism vs. Ism
flaxenwave.blogspot.com/2011/04/ism-vs-ism.html - Cached
15 Apr 2011 – This is why an architect must be a **good house occupant, but the Symbolists were poor** craftsmen. To build means to do battle with emptiness, ...

Figure 7: “Good house * * poor”

3.4 Exalead Search Engine

Exalead14 is a newer search engine that offers powerful search functions. It's major strength is that it offers several search operators that are not available with other major search engines. It also supports the NEAR operator. The syntax for such a query is word1 NEAR/n word2 where n is the word distance. It's major drawback is that it's search index is smaller than major search engines. It also does not have any API for making queries programatically.

Due to the availability of the NEAR operator, exalead search engine will be used as for calculating the semantic orientation of the phrases extracted.

14 http://www.exalead.com/search/

4.0 Experiment

In the experiment, the approach by (Turney 2001) for calculating semantic orientation of a phrase will be implemented with exalead search engine. The corpus of reviews are as follows: 50 (25 negative and 25 positive) automobile reviews as well as 50 phone reviews and 35 movie reviews.
The first step involves part-of-speech tagging. Approximately 3300 phrases were extracted and 6600 queries made to the search engine. Due to the expected fluctuations of internet search hit counts, the total hits for POOR and EXCELLENT in the database index has been set at 57678218 and 109080313 respectively. This are the values obtained on the first day of the querying process.

The queries were issued consistently to help minimize the possible search engine hits fluctuations.

The table below shows the number of reviews and an average of the phrases;

Domain	Number of Reviews	Average Phrases Per Review
Automobiles (Yes)	25	40.24
Automobiles (No)	25	24.04
Phones (Yes)	25	25.4
Phones (No)	25	12.68
Movies (No)	25	22.25
Movies (Yes)	10	19.9
Total	135	24.38

Table 5: The review corpus in numbers

The formula for calculating semantic orientation is then applied

$$SO(\text{phrase}) = \log \frac{\text{hits}(\text{phrase NEAR excellent})\ \text{hits}(\text{poor})}{\text{hits}(\text{phrase NEAR poor})\ \text{hits}(\text{excellent})}$$

where hits(poor) = 57678218 and hits(excellent) = 109080313.
Table 6 shows the results to an automobile review classified as negative. The average semantic orientation of the review is approximately -0.429. A negative value indicates not recommended or “thumbs down”. Therefore the review is not recommend.

no25	phrase NEAR/10 Poor phrase	phrase NEAR/10 excellent	Semantic Orientation
important points	1214	2236	-0.0114814288
large suvs	75	14	-1.005667769
new car	14373	23236	-0.0681205991
high car	235	76	-0.7669888113
high fuel	2393	1206	-0.5743299321
much gas	219	97	-0.6304069219
big risk	332	97	-0.8111008907
smaller cars	178	244	-0.1397647173
higher premium	624	362	-0.5132105605
normal sized	460	981	0.0521766344
more information	43232	243909	0.474687957
new trend	761	1044	-0.1394186994
major costs	135	77	-0.5205775846
recently purchased	1500	8429	0.4729502535
promptly changed	29	3	-1.2620112845
less average	38	18	-0.6012456328
really angers	31	8	-0.8650062481
constantly worry	102	33	-0.7668207732
most large	1177	751	-0.4718710671
		Average SO	-0.4288530566

Table 6: Semantic orientation calculation

Although the overall sentiment of the review is negative, two phrases are positively classified. These are “more information” and “normal sized”. Taken into context “It is, however, too deep for a normal sized coffee cup” - clearly the author of the reviews does not quite like the size of the coffee cup.

4.1 Summary of results

Domain of Review	Number of reviews	Correct	Incorrect	Percentage of Accuracy
Automobiles (Yes)	25	23	2	92%
Automobiles (No)	25	13	12	52%
Phones (Yes)	25	19	6	76%
Phones (No)	25	17	8	68%
Movies (No)	25	24	1	96%
Movies (Yes)	10	2	8	20%

Table 7: Result Summary

The movies (yes) has the worst rating with 20% percent accuracy. Although there are only 10 reviews of which only two classified correctly. It should be noted that the total phrases contained in those reviews is about 200. This means that about three quarters of the phrases were misclassified. It is expected that there will be a mix of positive and negative generally towards a review. However, relying on this algorithm for sentence-level classification will lead to very unreliable results. For instance, the following movie review (Twister):

Phrase: real plot
Sentence: "Twister was good but had no real plot and no one to simpithize with"

Phrase: amazing effects
Sentence: ". but twister had amazing effects and i was hoping so...."

Phrase: small earthquake
Sentence: "he worrys about a small earthquake enough to leave his daughter at home"

Phrase: Geologic event
Sentence: "its takes a geologic event to heat millions of gallons of water in 12 hours"

Phrase: few hours
Sentence: "a few hours later large amount of ash start to fall"
Phrase: large amount

Sentence: a few hours later large amount of ash start to fall

Phrase: volcanic eruption

Sentence: it starts . the volcanic eruption

cv178_tok-7453 phrase	NEAR/10 Poor phrase	NEAR/10 excellent	Semantic Orientation
real plot	139	91	-0.4607079492
amazing effects	10 3	02	1.2032724017
small earthquake	12	3	-0.8787945326
geologic event	1	2	0.0242954544
few hours	4343	7542	-0.0370378426
large amount	11541	15311	-0.1539744258
volcanic eruption	587	127	-0.9415689216
		Average SO	-0.1777879737

Table 8: Twister Movie review sentiment

From Table 8 above, it can be seen that the only phrase that has a positive semantic orientation is “amazing effects”. Rightly so, “amazing effects” will most likely be more strongly associated with excellent than poor. All the other phrases have negative semantic orientation. Earthquake and volcanic eruptions too will be more associated with poor and other undesirable words, perhaps negative is accurate. However, most of the phrases above are descriptions about the film rather than the sentiment of the reviewer towards the film itself.

The major contributing factor towards the misclassification of the review is the pattern for extracting two-word phrases from the corpus. Turney (2001) stated that an isolated adjective such as “unpredictable” taken out of context might be either negative or positive. In the example, “unpredictable steering” and “unpredictable plot” will have negative orientation and positive orientation respectively. However, the review above, the reviewer clearly expressed his overall sentiment towards the film with words like -

“ i liked this movie”, ...
“but it was not as great as i hoped ...”. .
it had excellent special effects .
. . it was good ! !"

The reason why this is misclassified is due to the fact that highlighted phrases do not conform to
the pattern as suggested i.e

1. Adjective and Noun
2. Adverb and Noun (third word can not be a noun)
3. Adverb and Verb

The phrase “i liked” is Noun and verb, “as great” adverb and adjective, “was good” is a verb and adjective.

From the above, phrases are not necessarily better than one words such as adjectives that show strong positive emotion.

5.0 Conclusion

Sentiment analysis is a challenging task. Although computers have more processing power nowadays, the problem is not about having more processing power but a natural language problem. This is even more evident in the movie review domain that makes the challenging task of determining semantic orientation of the text even more difficult.

From the experiment, it has shown that relying on two word phrases just to provide context is not necessarily enough to provide better accuracies. Other single words can carry sentiment equally and unambiguously and they should not be ignored.

The major drawback of the approach of implementing the PMI-IR algorithm to calculate semantic orientation with the Exalead search engine is that search engine hits are unreliable and can vary wildly from day to day. This is more evident with the bigger search engines.

Also the lack of programming API's that support proximity (NEAR) search has hampered the number of reviews tested. The 135 reviews however gave some insight into the limitation of the part of speech tags used. Also a detailed literature review has been carried in this project that has highlighted the state of the art. This is also one of the achievements of this project.

Bibliography

1. Turney, P.D Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. ACL '02 Proceedings of the 40th Annual Meeting on Association for Computational Linguistics.

2. Adrienne Lehrer. 1974. Semantic Fields and Lexical Structure. North Holland, Amsterdam and New York.

3. Turney, Peter, and Littman, Michael. 2003. Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems 21:315- 346

4. Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). Pages 79–86.

5. Stefan Evert Google Web 1T 5-Grams Made Easy (but not for the computer)

6. K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of WWW-03, 12th International Conference on the World Wide Web, pages 519–528, Budapest, HU,

7. G. Grefenstette, Y. Qu, J. G. Shanahan, and D. A. Evans. Coupling niche browsers and affect analysis for an opinion mining application. In Proceedings of RIAO-04, 7th International Conference on "Recherche d'Information Assist´e par Ordinateur", pages 186–194, Avignon, FR, 2004.

8. S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima. Mining product reputations on the Web. In Proceedings of KDD-02, 8th ACM International Conference on Knowledge Discovery and Data Mining, pages 341–349, Edmonton, CA, 2002. ACM Press.

9. T. Nasukawa and J. Yi. Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the K-CAP-03, 2nd International Conference on Knowledge Capture,pages 70–77, New York, US, 2003. ACM Press.

10. B. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of ACL-04, 42nd Meeting of the Association for Computational Linguistics, pages 271–278, Barcelona, ES, 2004.

11. B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of EMNLP-02, 7th Conference on Empirical Methods

12. Z. Fei, J. Liu, and G. Wu. Sentiment classification using phrase patterns. In Proceedings of CIT-04, 4th International Conference on Computer and Information Technology, pages 1147–1152, Wuhan, CN, 2004.

13. Moshe Koppel and Itai Shtrimberg. Good news or bad news? Let the market decide. In Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications, pages 86–88, 2004.

14. Brill, E. 1994. Some advances in transformation-based part of speech tagging. Proceedings of the Twelfth National Conference on Artificial Intelligence (pp. 722-727). Menlo Park, CA: AAAI Press.

15. Church, K.W., & Hanks, P. 1989. Word association norms, mutual information and lexicography. Proceedings of the 27th Annual Conference of the ACL (pp. 76-83). New Brunswick, NJ: ACL

16. comScore and Kelsey Group 2007. Online Consumer-Generated Reviews Have Significant Impact on Offline Purchase Behavior

17. Tong, R.M. 2001. An operational system for detecting and tracking opinions in online discussions. Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification (pp. 1- 6). New York, NY: ACM

18. Alison Huettner and Pero Subasic. 2000. Fuzzy typing for document management. In ACL 2000 Companion Volume: Tutorial Abstracts and Demonstration Notes, pages 26 – 27

19. Kushal Dave, Steve Lawrence, and David M. Pennock. 2003. Mining the peanut gallery:opinion extraction and semantic classification of product reviews. International World Wide Web Conference, page 519

20. Arun Meena and T. V. Prabhakar. 2007. Sentence level sentiment analysis in the

21. presence of conjuncts using linguistic analysis. In Proceedings of the 29th European conference on IR research (ECIR'07), Giambattista Amati, Claudio Carpineto, and Giovanni Romano (Eds.). Springer-Verlag, Berlin, Heidelberg, 573-580.

21. Rebecca Bruce and Janyce Wiebe. 2000. Recognizing subjectivity: A case study of manual tagging. Natural Language Engineering, 6(2).

22. J. Wiebe, R. Bruce, and T. O'Hara. 1999. Development and use of a gold standard data set for subjectivity classifications. In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL-99), pages 246–253, University of Maryland, June.

23. S. Baccianella, et al. (2010). 'SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining'. In N. C. Chair, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (eds.), Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta. European Language Resources Association (ELRA).

24. The Stanford Natural Language Processing Group (http://nlp.stanford.edu/software/lex- parser.shtml)

25. Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, Daniel M. Ogilvie, and associates. 1966. The General Inquirer: A Computer Approach to Content Analysis. The MIT Press.

26. Princeton University "About WordNet." WordNet. Princeton University. 2010. <http://wordnet.princeton.edu>

28. WordNet: An On-Line Lexical Database. http://www.cosgi.princeton.edu/~wn.

29.Fellbaum, C., D. Gross, and K. Miller. 1993. Adjectives in WordNet. http://www.cosgi.princeton.edu/~wn.

30. S. Kim and E. Hovy 2004. Determining the Sentiment of Opinions. In Proc. Of the International Conference on Computational Linguistics (COLING '04).

31. Riloff, E., Wiebe, J. 2003. “Learning Extraction Patterns for Subjective Expressions”,Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing(EMNLP), Japan, Sapporo, 2003

32. Hong Yu and Vasileios Hatzivassiloglou. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2003.

33. Vasileios Hatzivassiloglou, Judith Klavans, Melissa Holcombe, Regina Barzilay, Min-Yen Kan, and Kathleen McKeown. 2001. SIMFINDER: A flexible clustering tool for summarization. In Proceedings of the Work-shop on Summarization in NAACL-01.

34. Beth Levin. 1993. English Verb Classes and Alternations: A Preliminary Investigation. University of Chicago Press.

35. T. Ballmer and W. Brennenstuhl. 1981. Speech Act Classification: A Study in the Lexical Analysis of English Speech Activity Verbs. Springer-Verlag.

36. Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP).

37. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts", Proceedings of the ACL, 2004.

38. Taboada, M. and J. Grieve (2004). Analyzing Appraisal Automatically. American Association for Artificial Intelligence Spring Symposium on Exploring Attitude and Affect in Text. Stanford. March 2004. AAAI Technical Report SS-04-07. (pp.158-161).

39. Turney, P.D. 2001. Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. Proceedings of the Twelfth European Conference on Machine Learning (pp. 491-502). Berlin: Springer-Verlag.

40. Exalead Web Search Engine. 2011. [Online] Available at:<http://www.exalead.com/search> [Accessed 14 August 2011]

41. Google Web Search Engine. 2011. [Online] Available at:<http://www.google.com> [Accessed 14 August 2011] [Accessed 14 August 2011]

42. Yahoo! Web Search Engine. 2011. [Online] Available at:<http://www.yahoo.com>