

## Expression and discussion: Sentiment analysis based on the text of forum replies in the case of ‘Songling Lin’

HE Jingjie

### Abstract

Based on the heatedly discussed social incident - “six policemen in Harbin beat a young man to death”, this study concentrates on analyzing the text of relevant comments posted by Chinese netizens on the forum of “Mop”. In order to figure out how the Chinese netizens behave online when discussing prominent and controversial social issues, the author has utilized the method of sentiment analysis and divided the research object into two dimensions - the netizens’ individual expression and the responded-to discussion among them. By respectively examining the length, punctuation and opinion-topic relation of the forum replies and measuring their sentiment polarities, the results revealed that most netizens’ self-oriented replies tend to express negative, radical and extreme emotions toward the government and authorities with only a few words. During the process of online discussion, the relationship among participants is more likely to grow into antagonism. Moreover, the netizens involved in debates tend to behave in a more aggressive manner, thus most debates eventually deviate from the principal contradiction and turn into personal attacks.

### Keywords

Sentiment analysis, text of forum replies, netizens’ online behavior

### Author

He Jingjie is a teacher at the School of Presenting Art, Communication University of China. Email: hejingjie0823@163.com.

## 网络表达与公共讨论： 基于“林松龄事件”中论坛回帖文本的情感分析

贺敬杰

## 摘要

基于“林松龄”社会热点事件,本研究对中国网民在“猫扑”论坛中的回帖文本进行了分析。为了通过应用情感分析方法进一步明确中国网民针对争议性热点事件的线上行为特征,本研究将研究对象分成两个维度,分别是网民个人化表达行为及网民回复性讨论行为,并主要通过提取相应回帖文本的长度、标点、观点与话题的关系等信息以及测量回帖文本的情感极性,来分别得出个人化表达行为与回复性讨论行为的特征。研究发现,网民个人化的表达多趋向于以较简短的话语表达极其负面、激进、极端的情绪,且通常指向政府或相关权力机构;而就网民之间的讨论过程而言,绝大多数参与者之间都表现为对抗式关系;被卷入争论的网民则表现得更具攻击性,从而争论发展到最后多偏离原有话题而演变成相互之间的人身攻击。

## 关键词

情感分析、论坛回帖文本、网民线上行为

## 作者简介

贺敬杰,博士,中国传媒大学播音主持艺术学院教师。电邮:hejingjie0823@163.com。

## Introduction

When considering the meaning of the highly developed information and communication technology (ICT) or the advent of the new media, we can not only define it just as “artefacts or devices that enable or extend our abilities to communicate”, beyond that, it also involves “the communication activities or practices we engage in to develop and use these devices; and the social arrangements or organizations that form around the devices and practices”. (Lievrouw & Livingstone, 2006) From this perspective, it is necessary to analyze the relationship or interaction between ICT and society. As we all know, the development of the advanced technology makes the internet’s emphasis on the ethos of democracy and egalitarianism possible, due to which a large number of citizens begin to use the internet as their main speech channels. We can see the fact that only by clicking the mouse the citizens can access to various alternative information sources. That is to say, the citizens, nowadays, no longer serve as a passive role in receiving information from the conventional authoritative news institutions. On the contrary, they have become more initiative in taking part in diverse online activities, such as discussing with other participants about some celebrities or prominent social events by leaving their comments on some forums. These countless comments can indicate the citizens’ viewpoints and emotions. And the relationships among themselves formed during the process of the discussion, to some extent, may demonstrate the general habits or characteristics of the citizens’ online behaviors. In this study, based on the highly controversial

social events that “six policemen in Harbin beat a young man to death”, examination will not only focus on the characteristics of comments given by netizens but also on the relationships among them. The method of sentiment analysis will be applied to meet the research target.

### **Introduction to this event**

At 10 pm, on October 11th, 2008, a violent incident happened in front of a bar gate in the city of Harbin, which caused a person named “Songling Lin” beaten to death. This incident was first reported by an authoritative internet-based news agency and then its influence was largely amplified after countless citizens had read a post named “six policemen in Harbin beat a young man to death” on an online forum named “Mop”. Two days later, on October 13th, the relevant authority released a section of the monitoring video which had recorded almost the entire process of the incident. The video showed that at first the young man was just at odds with the policemen but then he went forward to attack them. After three times of defiant violent behaviors conducted by the young man, one policeman’s head was hit and injured. The policemen eventually became outrageous to fight back and beat the young man to death.

### **Literature review**

The prospects of democracy related to netizens’ online participation

When it comes to the prospects of democracy, arguments on netizens’ behavior are always associated with the possibility of netizens’ effective participation and the challenges or problems of netizens’ participation.

It is widely believed that the achievement of the well-developed information technology, especially the advent of the internet, results in the formation of people’s new communication and participation patterns. In the past, the traditional mass media technologies embedded in the concepts of the industrial-era social institution which is characterized as a hierarchical configuration. In this hierarchical construction, the large-scale production and distribution of news only derived from very few authoritative media centers, whereas the audiences only received the messages passively. That means the information just disseminated in one direction from the “top/center” of the hierarchy to the “bottom/periphery”. As in today’s communication environment, the network provides a wide and multicomponent interconnection in which countless points or nodes are embodied. The audiences, thus, are just like the nodes which can be either a producer/sender or a receiver of information.(Wessels,2010) A research conducted by Katz etc al. relating to the usage of the internet and its impact on social interaction shows that compared to the citizens offline, the internet-users are likely to possess wide social networks of friends and experience increasing exposure to information presented in diverse websites, all of which drives citizens to take part in more

formal or informal events or activities. (Katz, Rice & Aspden,2001)That is to say the internet seems to have the potential to promote the process of expanding citizens' participations in public affairs. Barney also argues that the internet which is equipped with networked logic has furnished countless forums for debates and makes the exchange of information and interactive communication among citizens or between citizens and their representatives possible.(Barney,2004) In this respect, the democratic process is enhanced due to the application of the internet.

Although the research above has indicated a positive potential for the netizens' effective online participation, it seems to be too optimistic to draw such an ideal democratic picture. Actually, the cyberspace indeed encounters potential problems and challenges in fulfilling the democratic promise, especially in achieving the effective and equal participation. They are anonymity and intolerance.

As an important characteristic of the internet, anonymity directly implies the reduction of social cues on citizens when they act online. In this context, people's fears of "isolation, humiliation, harming others, not being liked, or disapproval" would be reduced(Shane,2004)and once "their actions cannot be attributed to them personally, they tend to become less inhibited by social conventions and restraints". (Wallace,1999) The most obvious consequence is the frequently happened conflicts or quarrels. This is because the anonymity stimulates the worst within netizens by enabling them to escape the accountability of their activities. Just as Davis (1999) argues that, "Usenet political discussion tends to favor the loudest and most aggressive individuals". Some also argue that "anonymity - like the myth of Gyges's ring that makes the wearer invisible - leads inexorably to immoral and even illegal behavior". (Saco,2002)

A study conducted by Robinson, Neustadtl, and Kestnbaum to investigate the differences between American net users and non-users regarding their political tolerance attitudes found that net users are more likely to be tolerant independent of their demographic background, such as social status and the level of the education and income. (Robinson, Neustadtl & Kestnbaum,2004)However, it is not difficult to detect an obvious phenomenon that most netizens tend to be irritable and intolerant when interacting with someone with opposite position. Sunstein demonstrated this phenomenon as the trend of "group polarization", which means that the members of a like-minded group would possess a more extreme position when discussing about a social or political issue, and "If diverse groups are seeing and hearing quite different point of views, or focusing on quite different topics, mutual understanding might be difficult, and it might be increasingly hard for people to solve problems that society faces together".(Sunstein,2001) The consequence to this is the overflow of furious flame wars in which "a person would respond to an 'offensive' communication with a scathing reply, often replete with obscenities and personal denunciations". (Margolis

& Moreno-Riano,2009)

The influence of postmodern culture in forming the netizens' activities online

There are also some researchers who demonstrate the netizens' behavior by relating it to the postmodern culture. As Castells mentioned that the most important characteristic of the postmodern social culture is the phenomenon of deconstruction which directly brings about the fragmentation of society. (Castells,1996) Harvey also characterized the postmodernism as being equipped with "penchant for deconstruction bordering on nihilism". Thus, the formation of deconstruction turns out to be the key point in describing the netizens' online behavior. (Harvey,1990)

Owing to the tendency of the commercialization spreading to the field of the internet, the cyberspace is gradually embedded with the commercial principles which will result in the emergence of commoditization and thus make the internet more likely depend on the usage by mass audiences rather than the small-scale groups of elites. The fact that the established traditional mass communication institution has been replaced by the fragmentary high-tech based communication environment directly engenders "the erosion of cultural hierarchies" and "the questioning of epistemological foundations and metanarratives". (Smart,1993)In other words, it means the fragmentation of cultural and cognitive orders. (Chaney,2002) As Chaney argues, the word "fragmentation" in cultural context does not mean the culture is turning into less important but refers to the culture dominated by authority is experiencing more and more challenges from diverse perspectives. Consequently, in the context of postmodern society, there is a growing tendency to critique the authorities, challenge the previously existing social institutions and break the conventional values. These kinds of subversive and denying behavior conducted by netizens also bring about the vanishing of center-controlled consciousness and the diversification of the orientation on the value, which makes the evaluation criterion vague and uncertain. That is to say, the lack of certain foundations for culture or cognition has a chronic preference to produce uncertainty and engender reconfiguration. On the one hand, it is helpful for people to get rid of constraints from traditional concepts. On the other hand, it also gives rise to radical skepticism and nihilism which imply great potentialities of disorder in both social and cultural context. Just as the British sociologist Giddens argued in his theory of "Utopian realism", our world will end up as a true postmodernity characterized as comprising post-scarcity order.(Giddens,1990)

### **Hypothesis**

Based on the above-noted literature, this study attempts to test the following four hypotheses.

HP1: Due to the characters of online anonymity, most Chinese netizens tend to be aggressive, excessive, and radical in expressing their opinions toward this controversial social event.

HP2: Influenced by the logic of deconstruction in postmodern culture, the Chinese netizens' evaluation criterion toward this incident is vague and uncertain. Further to say, they just focus on accusing the authority rather than paying attention to criminal fact.

HP3: The dialogue usually occurs when netizens hold different views; that is to say, the conformity in both sides of conversation is scarce.

HP4: Apart from debates or exchange of information among netizens, the forum also breeds impatience and intolerance which make countless extreme and meaningless disputes or quarrels among Chinese netizens possible.

## Sample and Methodology

### Sampling

As mentioned before, the controversial news that six policemen in Harbin beat an undergraduate to death had been widely spread owing to a post that had been presented in the online forum named "Mop". In order to acquire enough relevant samples, this study focuses on the netizens' comments to the post that presented in that forum. Firstly, in order to find that post, several words like "policemen in Harbin fight a person to death, Mop" have been typed in the search engine of that forum. Then, among all of the search results, the needed post has been quickly located according to its exact releasing time (October 12, 2008). After clicking it and entering, one can see that the total number of page views is 31787 and replies are 6356. However, only 4500 of the total 6356 replies are available.

Secondly, due to the fact that not all of the available replies are effective and valuable for this study, it is necessary to do some screening work. There are four criteria to consider in filtering the invalid comments. First, if a comment contains only a picture or just some punctuation, it would be considered as invalid, such as the 109th floor's netizen comment- "...". Second, comments that have already been prohibited or deleted by the forum administrator were naturally excluded. Third, some netizens would repeatedly leave comments of the same content, within which only one comment can be identified as valid. Fourth, comments with no more than three words would be defined as invalid. Just as the 409th floor's netizen commented, "Passing by ~~~" (translated by researcher). After filtering 524 invalid samples, the final sample size is 3976.

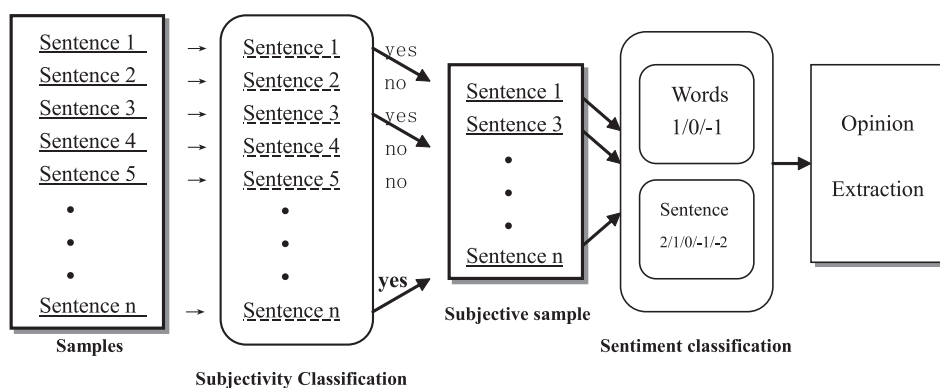
### Application of sentiment analysis

In this study, a quantitative method, sentiment analysis, will be used to measure and estimate the Chinese netizens online behavior. As to the development of the phrase sentiment analysis, it is

primordially used in analyzing market sentiment defined by Das and Chen in 2001 as the “automatic analysis of evaluative text and tracking of the predictive judgments”. (Pang & Lee,2008) That is to say, originally, the sentiment analysis played as a significant role in helping the online merchants get direct customers’ feedbacks toward their products so as to improve their products and adjust their sales strategies. Subsequently, countless studies with regard to the “sentiment analysis” begin to narrowly concentrate on the application of classifying comments’ polarity, be it positive or negative. These days the understanding toward this phrase has been extended to a broader extent covering the computation-based analysis of opinion, sentiment and subjectivity in both authentic and cyber text. Apart from its two main applications in business and politics, there are also some other possible applications of sentiment analysis, comprising the “online message sentiment filtering, E-mail sentiment classification, weblog author’s attitude analysis, sentiment web search engine and so forth”. (Tang, Tan & Cheng,2009)

The application of sentiment analysis has recently experienced a great increase in the domain of academic research. The notion of sentiment analysis can also be demonstrated as “sentiment detection” or “opinion mining”. It is a process of detecting “subjective information contained in texts, including viewpoint, fancy, attitude, sensibility, etc.” which provides a useful tool for market decision makers or social psychologists to figure out “whether a product review is positive or negative; what are the moods among Bloggers at that time; how the public reflect towards this political affair, etc.” (Tang, Tan & Cheng,2009) With respect to the methodological framework (chart 1) of sentiment analysis, it always involves “subjectivity classification”, “sentiment classification”, “opinion-topic relation extraction” and “participants’ relation extraction”. (Tang, Tan & Cheng,2009)

**Chart 1: The methodological framework**



### 1. Subjectivity classification

According to Wiebe, the notion of subjectivity in natural language means the individuals utilize the language to express views and estimation. (Wiebe,1994)It is on the opposite side compared to the objectivity which implies fact-based judgment without being influenced by the emotions or personal prejudices. The mission of the subjectivity classification is to discriminate the subjective sentences from the objective one. As Tang et al. said in their studies, the subjectivity classification is stated as follows: “ $S = \{s_1, \dots, s_n\}$ ” stands for “a set of sentences” in which “subjective sentences set  $S_s$ ” and “objective sentences set  $S_o$ ”. (Tang, Tan & Cheng,2009)

As to the specific approaches that are used to fulfill the work of subjectivity classification, there are two main approaches. One is making use of the correlation between the existence of adjectives or adverbs and sentence subjectivity to directly detect the adjectives or adverbs in sentences so as to tell the sentence is subjective or objective. This approach was first demonstrated by Hatzivassiloglou and Wiebe. (Hatzivassiloglou & Wiebe,2000) (Wiebe,1994; Yu & Hatzivassiloglou,2003)They tested the impacts of adjective orientation and sentiment value on sentence subjectivity. And then telling a given sentence is subjective or objective by just judging from the adjectives consisted in that sentence. Another approach is named Naive Bayes classifier which is a more accurate approach in distinguishing the subjective sentences from objective sentences compared to the former. It comprises “the counts of positive and negative words in the sentence, as well as counts of the polarities of sequences of semantically oriented words” (Tang, Tan & Cheng,2009)For instance, the “+ +” stands for two successive positively oriented words appearing in a sentence. Moreover, the factors to test are not only limited to adjectives, but also include the primary verbs and so on. Take the research conducted by Yu and Hatzivassiloglou (Yu & Hatzivassiloglou,2003) for example, this study was said to attain high accuracy with the rate of 97% by using the approach of Naive Bayes classifier.(Yu & Hatzivassiloglou,2003)

### 2. Sentiment classification

After extracting the subjective sentences, the following work is to value or classify the given sentences' sentiment. Sentiment classification encompasses two kinds of classification patterns, be it “binary sentiment classification and multi-class sentiment classification”. (Tang, Tan & Cheng,2009)As in binary sentiment classification, the sentiment is divided into two sides, respectively be positive and negative. Whereas in multi-class sentiment classification, the degree of the sentiment has been divided into five sections—strong positive (set 2), positive (set 1), neutral (set 0), negative (set -1), strong negative (set -2). The latter is usually applied to rank products or compare various individuals' online views or comments.



However, it needs to note that the notion of neutrality is different from that of the objectivity. The neutrality of opinion refers to the vagueness on whether the sentiment is positive or negative. It is also an expression of sentiment which may not contain factual information. However, the objective comments are far from relieving one's feeling but a must to present factual information.

### 3. Opinion-topic relation extraction

The individuals' opinions or comments are based on some specific topics. According to, "Opinion-topic relation refers to relationship between opinion expression (opinion-bearing words) and document topic (or feature of topic)". (Tang, Tan & Cheng, 2009) Once concentrating on extracting opinion-bearing words that have the potential to reflect sentiment polarity toward the given topic or feature of the topic, we can get a more precise research outcomes. (Rosario & Hearst, 2004) So it is necessary to analyze these two elements simultaneously rather than separate the comments from the topics and then mechanically consider the comments' sentiment orientation. For instance, the comment of "go read the book" (furnished by Bob Bland) is more likely to convey positive sentiment for book, nevertheless, when it responds to a movie, it turns out to be a negative comment. (Pang & Lee, 2008) Thus, exactly similar comments sometimes can imply opposite sentiments when mention to different objects.

### 4. Participants' relation extraction

A study on 100 responses in newsgroups toward three distinct controversial topics conducted by Agrawal et al. found that the relationship between two netizens in the "responded-to" network turns out to be more hostile (74% of the total responses are opponent compared to 7% are consentaneous). (Agrawal, Rajagopalan, Srikant & Xu, 2003) It is an interesting and gradually pervasive phenomenon in the internet, especially in politics related forums, online discussion boards or some blog posts, that the netizens express their opinions not only for the sake of responding to some specific controversial topics but also with the purpose of arguing with or even persuading the other participants. Nowadays, more netizens tend to focus on debating with others who possess different opinions toward a same topic rather than satisfied with merely expressing their own point of views. Worse still, in order to pursue the unique and unconventional personality, some netizens intentionally stand on the opposite side of other participants regardless of what their own opinions really are. Consequently, in this study, the Chinese participants' relation in the discussion about this controversial event has also been taken into consideration.

### Independent variables and coding process

Based on the methodological framework and a previous relative research in which six feature variables including "Frequency Counts", "Level of Subjectivity", "Length Features", "Punctuation",

“Polarity”, and “Referral” about the disputes in Weblog comments were set, (Mishne & Glance,2006) this survey sets eight variables.

#### Types of comments

In this study, to do the coding work, the first thing is to categorize the types of the netizens' comments toward a specific controversial topic. The comments are divided into four categories: personal-based comments, “responded-to” comments, citation comments and objective comments. The personal-based comments are kinds of comments that just contain the netizens' personal views or feelings toward the given topic no matter they are favorable or contradictory. The “responded-to” comments are associated with two individuals' conversation which implies the relationship between the two netizens, be opposed or coincident. And the opposite relationship would result in a furious debate or even obloquy. With regard to the citation comments, some netizens would just reference some topic-related news or information to express their opinions or feelings. Apart from all the three subjective kinds of comments above, the remaining comments are all classified as objective, for example, the comments totally present fact-based information without attaching any personal opinions or feelings.

#### Dates of comments

Because this event continued on October 13, 2008 when the relevant authority released the surveillance video that recorded the whole process, this study would divide the whole comments into two parts based on the comments' dates to analyze the difference of the netizens' attitudes. The first part of comments is given by netizens from October 12, 2008 to October 13, 2008. The second part is consisted of the comments given by netizens after October 14, 2008.

#### Length

Taking the length of comments into account as an important variable is to figure out the Chinese netizens' habits in leaving comments toward a controversial political related topic. Whether the comments are long or short can imply to some extent the participants' emotions. Just as Mishne and Glance observed in their research, the “disputative comments tend to be longer and appear in longer threads”.(Mishne & Glance,2006) In this study, the mean value of the comments' length would be calculated after using the “Microsoft Word” to count the comments' total number of words.

#### Comments' objects

As this study is based on a certain social event, therefore, in order to figure out how the netizens interact online toward this certain controversy, it is primary to consider the netizens' comments' object. The comments' objects in given sentences always turn out to be nouns. For instance, in the sentence of “it is

absolutely the dereliction of duty of the government and relevant departments that had made things worse”, the objects are “the government and relevant departments”.

#### Words' frequency counts and sentiment value

It is well understood that the words' frequency serves as a basic and popular variable set in content analysis works or in text classification tasks ].(Sebastiani,2002) In this study, owing to the majority of the adjectives and adverbs as well as the minority of nouns and verbs owning the semantic orientation, the coding work of counting words' frequency and valuing their sentiment level would focus on adjectives and adverbs. (Gamon & Aue,2005)And by finding out the words that are frequently used, we can know to some extent about most netizens' habits of expressing their personal emotions.

In addition, in order to rank the level of the sentiment expressed in netizens' comments, this study would make use of the SentiWordNet1 thesaurus which serves as a dictionary-based approach providing the words' sentiment annotations. According to Esuli and Sebastiani the SentiWordNet is a lexical resource based on the WordNet2 thesaurus which comprises textual representations of words. (Esuli & Sebastiani,2006)We could acquire almost all the words' sentiment mark in the forms of a triple of three sentiment values (p: #, o: #, n: #). “P” refers to positive value, “O” means words' objective or neutral sentiment value and “N” corresponds to words' negative value. Each sentiment values is in the range from 0 to 1 and the three must add up to 1 for each triple. Take the words of “beautiful” and “ugly” for example. The word “beautiful” = (p: 0.75, o: 0.25, n: 0) and “ugly” = (p: 0, o: 0.25, n: 0.75). In this study, the positive words would range from +0.3 to +1, the negative words would range from -1 to -0.3 and the words' sentiment would be defined as neutral if the sentiment value is between -0.3 to +0.3, such as the neutral word “similar” = (0.25, 0.75, 0) would be marked as “+0.25”. Two things need to be noted. First, as there are an increasing number of new words being used frequently by Chinese netizens, it is necessary to figure out the new words' exact meanings. For instance, the term “BS” is equal to the word “despise”. Second, when searching a word in SentiWordNet thesaurus, we will get more than one result, each of which is differentiated by the word's different part-of-speech pattern and contextual meaning. Consequently, in order to verify a word's sentiment value, its part-of-speech pattern and contextual meaning in a certain sentence have to be considered simultaneously.

#### Sentences' polarities

As it is stated before, the sentences' sentiment orientation can be deduced by valuing the sentiment of the words included in these sentences. In this study, due to the possibility for some comments comprising more than two sentiment oriented words, the sentences' sentiment orientation should be divided into five

hierarchies: strong positive, positive, neutral, negative, and strong negative. Based on the classification of the words' sentiment value, the strong positive sentences= " $> +1.5$ ", positive sentences=  $[+0.5, +1.5]$ , neutral sentences=  $[-0.5, +0.5]$ , negative sentences=  $[-1.5, -0.5]$ , strong negative sentences= " $< -1.5$ ".

However, when analyzing the sentences' polarities, there are several extra factors that should be considered. First, in order to make the study outcomes more precise, it is important to deal with the negation condition occurring in sentences. According to the previous research, the words such as "not", "no", "don't" or "hardly" or so can directly demonstrate the negation. Therefore, when doing the coding work on a sentence like "I don't like this book", the token "like" is transformed into the new token "like-NOT". (Das & Chen, 2001) Second, it is also necessary to consider the conjunctions between adjectives within one sentence. For instance, the word "and" always joints two adjectives of the same sentiment orientation, whereas, the word "but" often joints two adjectives with the opposite sentiment orientation. Lastly, within Chinese linguistic context, people get used to expressing their opinions or views in a relatively implicit or even ironic manner. In this condition, the netizens' negative comments always contain none of the apparently negative words. For example, in the sentence of "If you are reading this because it is your darling fragrance, please wears it at home exclusively, and tape the windows shut" (review by Luca Turin and Tania Sanchez of the Givenchy perfume Amarige, in *Perfumes: The Guide*, Viking 2008, cited by Pang and Lee, 2008) (Pang & Lee, 2008)

#### Standpoints

This study is based on the analysis of the netizens' responses toward the event regarding six policemen beat an undergraduate to death. Thus, it is necessary to analyze the netizens' standpoints rather than just examine every comment's sentiment orientation. The total netizens' comments would be divided into four different standpoints: stand in the policemen's position, neutral, stand in position of the dead, and the others.

#### Punctuation

As is well known that the applications of certain punctuation, such as the interrogation mark (?), the full stop mark (.), and the exclamation mark (!), to some extent would symbolize different levels of the sentences' sentiments. According to the previous research conducted by Sahami et al., counting the frequency of diverse punctuation is proven to be effective for the tasks of text classification. (Sahami, Dumais, Heckerman & Horvitz, 1998)

## Results and Analysis

### General characteristics of the Chinese netizens' comments

Subjective comments outnumber objective comments and the “responded-to” comments are limited

A glance at Chart 2 reveals that the whole valid netizens' comments had been divided into four parts: personal-based comments, responded-to comments, citation comments and objective comments. Within all of these four kinds of comments, the personal-based comments dominated with the frequency of 2986 accounted for 75 percent of the total valid comments, standing in marked contrast with the citation comments (125, 3 percent) which always emerge as some celebrities' sayings, such as the 445th floor and the 1477th floor netizens both cited the Chinese famous writer—Lu Xun's verse and article to express their concerns about the current reality of Chinese society, or some widely concerned social events that had already happened. For example, the 911th floor cited two controversial social news (“milk scandal” in Chinese “毒牛奶” and “Zhou tiger” in Chinese “周正农 ‘华南虎’”) to emphasis his/her attitude toward the authorities. After the personal-based comments, the responded-to and objective comments came in turn with 485 (12 percent) and 380 (10 percent).

**Chart 2: Types of comments**

Types of comments	Frequency	Rates
Personal-based comments	2986	0.75
Responded-to comments	485	0.12
Citation comments	125	0.03
Objective comments	380	0.10
Total	3976	1

Considering all the first three kinds of comments as subjective comments, it is clear that the number of subjective comments (3596, 90 percent) is much more than that of the objective comments. It means that the majority of Chinese netizens tend to actively express their opinions and show their real emotions by leaving comments on the forum named Mop.

However, what beyond expectations is the small number of “responded-to” comments, which implies that the interaction among diverse netizens toward this event in forum is limited. Possibly because the forum is considered as a relatively closed cyberspace compared to the so-called social network site (Facebook, Sina Microblog and so on), in which the one-to-many rather than one-to-one modes of communication together with the lack of instant notifications on “responded-to” comments make a sustained interaction or dialogue impossible. That is to say, the forum as one kind of cyberspace is far

from being a “contested terrain” that is characterized by countless disputes or quarrels.(Dahlberg & Siapera,2007)

The majority of the comments contain only a few words

The length of netizens’ comments would demonstrate to some extent the habits of the Chinese netizens in responding to controversial social news. As every comment differs greatly in length (in this study, the minimum of comments’ length is 5 words and the longest sentence contains more than hundreds of words), therefore, it is necessary to set a mean value in sentences’ length. After copying all the valid samples to the Microsoft Word file, the total number of the whole given samples’ words would be worked out (115604 words), and then the mean value has been figured out (29 words). Given is a table (Chart 3) with regard to the netizens’ comments’ length. Based on the mean value, it is apparent that the comments that contain less than 29 words are much more than the comments with more than 29 words, respectively being 2612 (66 percent) and 1364 (34 percent). That means the majority of the netizens’ comments are short.

**Chart 3: Length**

Comments’ length	Frequency	Rates
More than 29 words	1364	0.34
Less than 29 words	2612	0.66
Total	3976	1

In this politically sensitive issue, we find that most netizens just used several adjective words or even curse words to abuse the related authorities rather than calmly organize their minds and logically present their views, such as the 18th floor netizen said that “Everybody, let’s overturn the police!” and the 914th floor netizen leaved a comment as “What a fucking society!” (Translated by researcher) Therefore, we can draw a conclusion that the fact that a multitude of netizens’ comments are short implies that the netizens tend to be impatient and easily become unreasonable or even infuriated toward a newly happened politically sensitive issue. In addition, there are few short comments showing the netizens’ feelings of uncertainty and doubt toward the given issue which indicates the netizens’ rationalities. For example, the 45th floor netizen leaved a reply as “Continuing to pay close attention to this issue” and the 103rd floor netizen said that “Is it true?” (Translated by researcher)

As for the comments that contain more than 29 words, this study finds that the total number of the relatively long comments is small and the long sentences always locate in the “responded-to comments” and the “citation comments”. The former means that when the netizens argue with or express support for

other netizens, they are willing to leave long comments. And the latter demonstrates that the rationality and dispassion in netizens are more likely to engender long comments. In addition, one particularly interesting fact worth mentioning is that there are several long comments just repeating the same sentence or comprising numerous repeated words, which also indicates the netizens’ impatience and their improper purposes to catch other netizens attentions, such as the 1647th floor netizen repeated a sentence for at least a hundred times.

The authorities always make an easy target for critics

After examining the objects of netizens’ comments and the changeableness of netizens’ standpoint, this study finds that the authorities are always destined to be condemned no matter whether they are innocent or not.

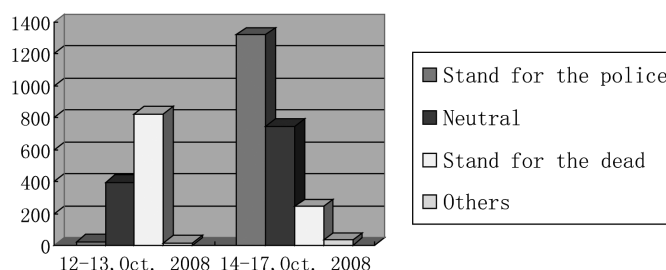
Judging from Chart 4, netizens in this controversial event mainly focus on commenting “The Police”, “The Dead” and “The Chinese Government”. The first two objects are highlighted by this event so that it is natural for them to be mentioned frequently. However, as for the object “The Chinese government”, there seems to be no direct connections with this event but it obtains a high frequency of being mentioned. On the contrary, there are only 3 percent of the whole valid comments referring to the words as “Fight or Assault”. Moreover, as to the sentiment orientations in relation to the authorities, it always turns out to be intensively negative. For example, the 1159th floor netizen said in his/her comment that “The police as ruffians (-0.875) are fostered by the extremely (-0.125) evil (-0.875) Chinese government”. (Translated by researcher) According to Chart 6, the polarity of this comment has been evaluated as “strong negative sentences” with the total sentiment value at “-1.875”. That is to say, the Chinese netizens mostly pay attention to accusing the Chinese government but fail to explore more details or facts and also have no patience to draw lessons from the incident itself.

Chart 4: Comments’ objects

Comments’ objects	Frequency	Rates	Total
The Police	1926	0.54	3596(1)
The Dead	1707	0.47	
Chinese Government	876	0.24	
Northeasterner of China	802	0.22	
Sports College Students	523	0.15	
Other Netizens	485	0.13	
Socialist System	171	0.05	
Fight or Assault	120	0.03	

In terms of the netizens' standpoint, in order to find out changes that have happened on it, the coding work comprises noting the dates of every valid comment. The number of the comments that given by netizens from the date of 12, October to 13, October of 2008 is 1314, and the data of the comments leaved by netizens from 14, October to 17, October, 2008 (the latest one) is 2662.

Chart 5: Standpoint



It is apparent by glancing at Chart 5 that in two connected periods of time, the netizens' standpoints experienced a tremendous transformation from accusing the misdeed of the police (symbol of authority) to interceding for the police and meanwhile blaming the dead for his unruly and unscrupulous behavior (the dead is said to be a "guan er dai", "官二代" in Chinese meaning the second generation of government officials). This kind of sudden change from condemning one authoritative image (the police) to another authoritative image (the "guan er da") also directly reveals the Chinese netizens' features of being blind, fickle and hasty rather than speaking and acting cautiously before knowing more about what the facts are. This also implies an instant and frequent communication environment in which the fast-paced cultural consumption is favored. Netizens cater for this rapid cyber culture, thus naturally become increasingly thoughtless. In this respect, it is not surprising for Chinese netizens being incited or manipulated by speeches with malicious motives.

We can explain this pervasive phenomenon from two aspects. Firstly, with the establishment of the postmodern environment in recent Chinese society, the appealing to deconstruction gradually engenders the fragmentation of the whole society. The most significant outcomes are the frequent emphasis on individualization and also the fading of the stress in authorities' power. People who live in such postmodern society, thus, become uncertain and vague in evaluation criterion, and are just keen on criticizing even accusing the authorities in an unreasonable manner. Especially in the relatively open and free cyber context, this kind of behavioral tendency has been intensified and amplified. Secondly, this phenomenon to some extent results from the development of the populism. Just as Chaney argues that nowadays within



the continually changed society and everybody's life, an emerging trend named popular democratization experiences a great extension, which makes the populism a dominant role among the public discourse. (Chaney,2002)However, populism may not bring about any popular authentic emancipation but result in more contradictions between individuals and authorities. Within the Chinese internet-based context, populism is developing to an extreme degree which tempts the majority of Chinese netizens to become extremely unreasonable in blindly condemning the authorities.

Apart from numerous aggressive, excessive, and radical comments, there are lots of negative or critical comments expressed in a relatively implicit manner.

After studying all the subjective comments, we can find that the majority of the netizens' comments turn out to be aggressive, excessive and radical. It can be demonstrated from three aspects - the hotspot words and their sentiment value (Chart 6), the classification of comments' polarities (Chart 7), and the usage of punctuations (Chart 8) in netizens' comments.

**Chart 6: Hotspot words' sentiment value, sorted by popularity**

Words	Property of words	Sentiment value
Gloomy (黑暗)	Adjective	-0.875
Hell (该死)	Noun	-0.625
Harmonious (和谐)	Adjective	+0.375
Damn (谴责)	Verb	-0.5
Fuck (操)	Verb	0
Brutal (野蛮的)	Adjective	-0.625
(son of) Bitch (婊子)	Noun	-0.75
Tragic (悲剧的)	Adjective	-0.625
Severely (严重地)	Adverb	-0.75
Ruffian (流氓)	Noun	-0.875
Rubbish (垃圾)	Noun	-0.125
(to be) Condemn (判刑)	Verb	-0.375
Shame (耻辱)	Noun	-0.75
Resentful (愤怒的)	Adjective	-0.875
Corrupt (腐败的)	Adjective	-0.875
Curse (诅咒)	Verb	-0.5
Salute (致敬、敬礼)	Noun	+0.375
Ugly (丑陋)	Adjective	-0.75
Apathetic (冷漠、冷血)	Adjective	-0.625
Extremely (极其地)	Adverb	-0.125

Chart 7: Sentences' polarities

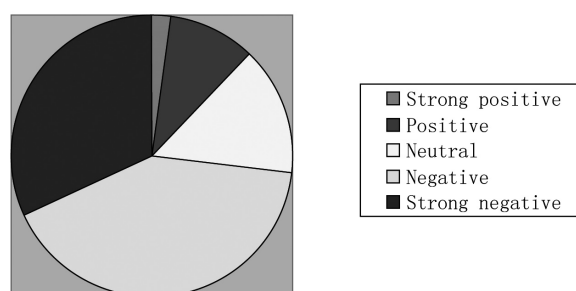
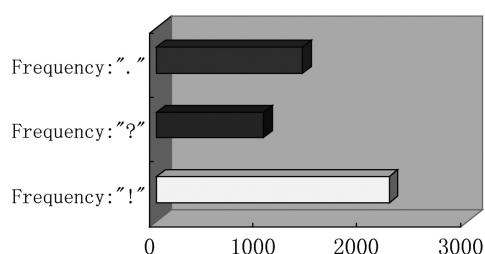


Chart 8: Punctuation



Regarding this issue, as to the words that the netizens frequently use to express their feelings, it is the adjectives and the nouns that respectively stand in the first and second place in the ranking list of the frequency of being mentioned, such as the adjectives as gloomy, brutal and the nouns as hell, (son of) bitch, rubbish. All of these words are sharing the same characters as negative words (evaluated within the range from -1 to -0.3) which imply the critical features of the netizens. Moreover, some of these words are valued as extremely negative words which indicate the aggressive and excessive characters of the Chinese netizens' comments. As for the comments' polarities, the data can also best demonstrate the netizens tendency of being radical and aggressive in responding to this issue or other participants. 41 percent negative sentences together with the 32 percent strong negative sentences (contain more than two negative oriented key words) dominate the netizens comments. Lastly, the netizens' frequent use of using the exclamation mark ("!") has illustrated the irritable aspect of the netizens' emotions to some degree. One special example that should be mentioned is the 1386th floor comment which contains 43 exclamation marks in just two sentences.

Just as Steiner said that "On the internet, no one knows you're a dog". (Steiner,1993)This kind of anonymity can best demonstrate why the Chinese netizens tend to be aggressive and radical by using

various negative words and numerous exclamation marks. Due to the obscureness or absence of every person's authentic online identities, the netizens would take part in various activities at will without considering whether their conducts are proper or not. Needing not to worry about undertaking the responsibilities for their inappropriate comments makes the netizens always blind to talk or dispute no matter it is right or wrong. In this study, we can find that the anonymous comments have been measured with more radical and aggressive sentiment orientation compared to the comments revealing the reviewers' identities.

In addition, also noteworthy in this study is the finding that not all the netizens are radical or aggressive when expressing their negative or critical views. There are still a number of Chinese netizens expressing their discontentedness in a relatively implicit manner which is defined as the irony. It may be influenced by the Chinese traditional culture in guiding people to be always implicit and euphemistic rather than to be impertinent or savage. The most typical example within this sensitive issue is that the words "harmonious" (positive word with sentiment value at +0.375) and "salute" (positive word with sentiment value at +0.375) has been frequently mentioned by the Chinese netizens in satirizing the misdeed of the police.

#### Discussion among diverse participants

In order to figure out the exact relation among Chinese netizens when they are discussing one hot social issue, it is pivotal to extract and then analyze the "responded-to" comments. This kind of comments refers to more than two netizens discussing an already given topic. In this study, although the dialogues among netizens in forum are limited, it is still feasible to extract some typical characters of the discussions online by synthetically examining the types, length, opinion-topic relation and sentiment orientations of "responded-to" comments.

Relationship in most cases will grow into antagonism when interaction happened among participants

Within the process of one discussion, the relationship between every two participants is forming, being either mutually exclusive or accordant. Thus, the types of the "responded-to" comments can be divided into two parts. Based on the data above, we have examined the overall 485 "responded-to" comments and then find that the proportion of the comments which indicate the participants' opposed relationships is largely exceeding that of the comments that reveal the coincident relationships among the netizens, with 384 and 101 respectively. That is to say, the majority of the conversations among netizens are defined as disputes. Imagining that if not for the limitations generated by technology platform, the forum is more like a noisy market full of wrangles rather than a so-called "market place of ideas".

When being involved in a dispute, Chinese netizens tend to be more positive or even aggressive

The length and sentiment polarities of the “responded-to” comments, to some extent, can demonstrate the netizens’ psychological state. We have copied all the 485 “responded-to” comments to the Microsoft word to count the total number of words and then work out the mean value of the “responded-to” comment’s length - 38 words - with the minimum length of 26 words. Compared to the overall comments’ length mean value of 29 words, the “responded-to” comments turn out to be generally longer than the other types of the netizens’ comments, which implies that the netizens have more desires to express their views when taking part in discussions, especially in disputes. A noteworthy fact that most of the contestable “responded-to” comments’ length is longer than that of the coincident ones, further indicates that the participants are willing to spend more time typing in a good many words to express their opinions when they argue with participants who stand in different or even opposite positions.

As to the sentiment polarities, within all the valid comments, the contestable “responded-to” comments (disputes) turn out to possess more negative and furious tones than “personal-based” comments. We can see the majority of disputes online are always closely associated with the hostile and offensive manner to abuse and calumniate the dissenters. We call this “intolerance”, which resulted from the wide spread of individualism (expansion and strengthening on the sense of egocentricity) and the effect of group consciousness. As netizens are usually willing to approach the views that hold the same orientation as their own, a group that consists of individuals with the same views is formed, in which the self-absorbed individuals would become narrowly focus on what they are convinced of. In this case, the netizens turn out to be more extreme and aggressive in persuading the others standing in different or opposite positions. This phenomenon is also termed as “group polarization”. Just as Sunstein argued that the internet is a “breeding ground for extremism” for like-minded netizens are mainly discussing among themselves but having limited time to listen to the differences. (Sunstein,2001) This kind of extremism would directly lead to not only the disputes but also the “flame wars”.

The disputes will easily deviate from principal contradiction and turn into personal attacks

This study also examined the features concerning the objects of the “responded-to” comments. As the study is based on a well-known news event regarding six policemen in Harbin beat a young man to death, the given topic is the “crime” while “the police” and “the dead” are features of the topic. It is natural for the netizens to leave comments including the objects of “the police” or “the dead”. However, after examining the objects of the “responded-to” comments, we find that most of the “responded-to” comments are assailing the “Northeasterner of China” with contemptuous language, which totally deviate

from the original topic. For example, the 213th floor netizen said that “the northeasterners of China are good at disturbing (-0.375) the social order and always turn out to be violent (-0.625), so it is good to kill (-0.5) more of them.” (Strong negative sentences with sentiment value at -1.5) (Translated by researcher)

Besides, there are also a lot of drastic disputes that concentrate on making personal abuse, which plays a negative role as disturbance in distracting the participants’ attention from the essential part of one event to its insignificant aspect.

## Conclusion

As the usage of the internet by netizens becomes more closely integrated with their daily lives, the ways in which the netizens interact with others and take part in various activities in cyber context are gradually considered as a kind of impetus that is associated with the progress of society and culture. We can see the fact that nowadays an increasing number of people have equal and diverse access to express their views, opinions and emotions when participating in some prominent social events. This kind of participation to some extent would enrich and activate the public opinion and thus produce more positive power in pushing forward the cyber democratic process. However, if we conduct a profound analysis on the phenomenon of the pervasive “free speech” in Chinese cyber context, we would find that there are some challenges and problems that make the prospect of the free speech even the cyber democracy more uncertain.

Based on the incident of “six policemen in Harbin beat a young man to death”, this study concentrates on analyzing the comments given by Chinese netizens on the forum of “Mop” by utilizing the method of sentiment analysis. Within the process of coding, this research set eight variables. And after gathering plenty of valid data, this study has explored the question of how the Chinese netizens behave online when discussing influential and controversial social incidents from several aspects, such as to consider the general characteristics of Chinese netizens’ comments by extracting opinion-topic relation and measuring the comments’ sentiment polarities, as well as to figure out the relationships among diverse internet-based participants by extracting the features of “responded-to” comments.

### Findings

◎The majority of the comments given by Chinese netizens on the forum turn out to be short which means Chinese netizens are too impatient to discuss this controversial event in a proper way. However, things go different when it happens to responded-to comments, most Chinese netizens will be generous with their time and words to argue with others fiercely and aggressively.

◎Most of the Chinese netizens' comments presented on the forum "Mop" are aggressive, excessive, and radical. We can find that the number of subjective comments is more than that of the objective comments which means the Chinese netizens are more likely to directly express their opinions and emotions rather than present fact-based information. What's more, netizens tend to utilize numerous negative adjectives, extremely uncivil nouns and exclamation marks to express their feelings or emotions toward one controversial social issue. Therefore, HP1 was supported. But there is an exceptional situation that some Chinese netizens are willing to express their negative or critical views in a relatively implicit manner, which defined as the situation of irony.

◎No matter what the fact is, most Chinese netizens' negative comments are concentrating in critiquing even accusing the authorities. The sudden change in netizens' standpoints from condemning one authoritative image (the police) to curse another authoritative image (the children of government officials) reveals the uncertainty and vagueness in Chinese netizens' evaluation criteria, which leads to blindness, fickleness and indiscreetness in Chinese netizens' online behavior. Thus, HP2 was supported.

◎Due to the one-to-many modes of communication and the lack of instant notifications on "responded-to" comments, the interaction or disputes among diverse netizens on the forum is limited. This result is not consistent with the HP4. But if we merely pay attention to Chinese netizens' online interactions, we can also draw some meaningful conclusions by analyzing the limited "responded-to" comments. First of all, the dialogues among netizens always indicate their antagonistic relationship, which verifies the HP3. Second, the Chinese netizens who are involved in disputes or quarrels turn out to be more unreasonable and intolerant. Moreover, the disputes always deviate from the given topic and eventually grow into personal attacks. These two findings partially support the HP4.

#### Limitations

This research just focuses on analyzing participants in forum of "Mop" who might merely represent only a small part of the overall Chinese netizens. So the sample size is small and the scope of the sampling is limited. In addition, owing to the Chinese language's complex and subjective features, there may be some deviations or misunderstanding happened in the process of coding work toward the whole valid comment samples. Last but most important, this study is conducted on the basis of online forum which is different from other forms of cyberspace such as social network sites, chat groups in instant messaging tools and so on. Shaped by certain cyberspace's characteristics, the netizens' behavior would exhibit different features. Therefore, future studies can extract and analyze the Chinese netizens' characteristics of expression and interaction based on different forms of cyberspace, if possible, the netizens' basic personal

information such as age, educational background should also be considered.

(责任编辑:王硕)

### 引用文献 [Reference]

- Agrawal, R., Rajagopalan, S., Srikant, R. & Xu, Y. (Mining newsgroups using networks arising from social behavior. WWW. Budapest, Hungary.
- Barney, D. (2004). *The network society*. Cambridge: Polity Press.
- Castells, M. (1996). *The rise of the network society*. Malden, MA: Blackwell Publishing.
- Chaney, D. (2002). *Cultural change and everyday life*. Basingstoke: Palgrave.
- Dahlberg, L. & Siapera, E. (eds.) (2007). *Radical democracy and the internet: Interrogating theory and practice*. Basingstoke: Palgrave Macmillan.
- Das, S. & Chen, M. (Yahoo! For Amazon: Extracting market sentiment from stock message boards. Asia Pacific Finance Association Annual Conference. Bangkok.
- Esuli, A. & Sebastiani, F. (SentiWordNet: A publicly available lexical resource for opinion mining. Language Resources and Evaluation. Genoa, Italy.
- Gamon, M. & Aue, A. (Automatic identification of sentiment vocabulary exploiting low association with known sentiment terms. the ACL workshop on feature engineering for machine learning in NLP. Ann Arbor.
- Giddens, A. (1990). *The consequences of modernity*. Cambridge: Polity.
- Harvey, D. (1990). *The condition of postmodernity: An enquiry into the origins of cultural change*. Oxford: Blackwell Publishing Ltd.
- Hatzivassiloglou, V. & Wiebe, J. (Effects of adjective orientation and gradability on sentence subjectivity. the International Conference on Computational Linguistics (COLING). Saarbrücken, Germany.
- Katz, J. E., Rice, R. E. & Aspdén, P. (2001). The Internet, 1995-2000: Access, Civic Involvement, and Social Interaction. *American Behavioral Scientist*, (45), 3.
- Lievrouw, L. & Livingstone, S. (2006). *The Handbook of New Media (student edition/2nd)*. London: Sage.
- Margolis, M. & Moreno-Riano, G. (2009). *The prospect of internet democracy*. England: Ashgate Publishing Limited.
- Mishne, G. & Glance, N. (Leave a reply: An analysis of weblog comments. Third annual workshop on the Weblogging ecosystem. Edinburgh, Scotland.
- Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 1 (2), 1-135.
- Robinson, J., Neustadtl, A. & Kestnbaum, M. (2004). Technology and tolerance: Public opinion differences among internet users and nonusers. In Howard & Jones, *Society online: The*

- Internet in context*. London: Sage.
- Rosario, B. & Hearst, M. A. (Classifying semantic relations in bioscience text. ACL' 2004. Barcelona, Spain.
- Saco, D. (2002). *Cybering democracy: Public space and the internet*. Minneapolis: University of Minnesota Press.
- Sahami, M., Dumais, S., Heckerman, D. & Horvitz, E. (A bayesian approach to filtering junk E-mail. Learning for Text Categorization. Madison, Wisconsin.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34 (1), 1-47.
- Shane, M. (2004). *Democracy online: The prospects for political renewal through the internet*. New York and London: Routledge.
- Smart, B. (1993). *Postmodernity*. USA and Canada: Routledge.
- Steiner, P. On the internet, no one knows you're a dog [J]. *New Yorker*, 1993(5 July).
- Sunstein, C. R. (2001). *Republic.com*. Princeton: Princeton University Press.
- Tang, H., Tan, S. & Cheng, X. (2009). A survey on sentiment detection of reviews. *Expert Systems with Applications*, (36).
- Wallace, P. (1999). *The psychology of the internet*. Cambridge: Cambridge University Press.
- Wessels, B. (2010). *Understanding the internet: A Socio-Cultural perspective*. UK: Palgrave Macmillan.
- Wiebe, J. (1994). Tracking point of view in narrative. *Computational Linguistics*, 20 (2), 233-287.
- Yu, H. & Hatzivassiloglou, V. (Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. the Conference on Empirical Methods in Natural Language Processing (EMNLP). Sapporo, Japan.