Words-of-Wisdom Search System based on Positive Negative Degree

Koichi Takaoka, Akiyo Nadamoto

***Abstract*—With the rapid advance of the Internet, anyone can obtain needed information easily. Nevertheless, few systems extract and present information that is suitable for a user's sentiment. In this paper, we present a system that instills a positive mood in users by extracting and presenting positive words-of-wisdom based on calculations of the user's sentiment. As described in this paper, as a first step of the sentiment search system, we particularly examine a Positive–Negative (P/N) value that is calculated for a user. In particular, we propose a method of calculating a word P/N value for words-of-wisdom, and propose the word-of-wisdom P/N value based on it. Then, we create a word-of-wisdom database based on our word-of-wisdom P/N value. In our proposed system, first a user inputs their current mood and desired mood. Then the system calculates the user's P/N value. The system subsequently extracts a word-of-wisdom which has a similar P/N value to that of the user's from the word-of-wisdom database. The authors conducted a user experiment to measure the availability of the proposed method.**

***Index Terms*—Sentiment extraction, Words-of-Wisdom, Positive/Negative**

I. Introduction

Recently, with the rapid development of the Internet, there is not only the general web, but also numerous information sharing services such as social network services, and word-of-mouth for online shopping. Consequently, with much related information on the Internet, it is difficult for users to extract the information that they seek. To solve that problem, many studies have examined extraction of sentiment information from web contents. For example, Hatzivassiloglou et al.[1] investigate the clustering of adjectives as either Positive or Negative. Leskovec et al.[2] examine such clustering for Positive or Negative in relation to Social Network Services. For example, when a user becomes pleasant, it is beneficial for that user to recommend pleasant information from mass information related to the Internet. They calculate the sentiment of content by extracting the sentiment words and quantifying them. It is beneficial for users to present selected information using the sentiment of contents.

People sometimes read words-of-wisdom when we feel melancholy or depressed. Words of wisdom might consist of encouraging phrases or text intended to cheer up someone else: they are called words-of-wisdom. There are many words-of-wisdom available on the Internet. For example, there are over 10,000 separately identifiable words-of-wisdom on a web site. Nevertheless, is difficult to extract the best word-of-wisdom for a particular user. We consider that it is convenient for users to extract word-of-wisdom on the Internet using sentiment measurement.

Manuscript received December 22, 2010. This work was supported in part by a Grant-in-Aid for the Information Explosion Project (Number: 21013044).

Koichi Takaoka is with the Konan University Graduate School, Kobe , Japan, e-mail: (tmk_kouichi@yahoo.co.jp).

Akiyo Nadamoto is with the Konan University, Kobe, Japan,e-mail: (nadamoto@konan-u.ac.jp).

As described in this paper, we propose a system that extracts a word-of-wisdom from the Internet using the user sentiment and characteristics of the word-of-wisdom. The flow of extracting word-of-wisdom can be represented as follows:

1) A user inputs information about the current mood and the desired mood.
2) We calculate a user's positive value from a user's current mood and desired mood.
3) We quantify the expression of word-of-wisdom, and calculate Positive/Negative(P/N) value. In this paper, we designate the value as "Word-of-wisdom P/N value". Our Word P/N Value is based on Word P/N Value.
4) We extract the words-of-wisdom that a user wants using a user's positive value and P/N value.

This paper is organized as follows: Section 2 describes related work. Section 3 explains the calculating user's positive value. Section 4 presents a description of the word P/N value. Section 5 presents a description of the word-of-wisdom P/N value. Section 6 presents a prototype system. Section 7 discusses results of experiments conducted using our system, and Section 8 presents the conclusions of our study.

II. Related Work

Recently, studies of sentiment expression extraction are active. Various sentiment models have been proposed. Multi-dimensional sentiment vectors are proposed in various sentiment models. For example, the model of Plutchik[3], [4] is a typical sentiment model. The model of Plutchik[3], [4] includes clustering of four-dimensional sentiment vectors based on eight emotions: "$acceptance''$ $\Leftrightarrow$ "$disgust''$, "$anticipation''$ $\Leftrightarrow$ "$surprise''$, "joy'' $\Leftrightarrow$ "$sadness''$, "$anger''$ $\Leftrightarrow$ "$fear''$.

Kumamoto et al.[5] specifically examine sentiments related to multi-dimensional emotions and then extract emotion from web news. The emotion model comprises two-dimensional sentiment vectors ("$sadglad''$$and$"$angry-pleased''$.), which are between 0 and 1. Their hypothesis is "A news article that includes emotion word 'e' represents the emotion. Their system was designed using co-occurrence between the word in the news article and emotion words.

Kawai et al.[6] classify news articles extracted from multiple news sites using the user profile. Their system is called MPV Plus, which presents a user's interests. In fact, MPV Plus extracts interest words and sentiment vector from the user's browsing history. They propose four-dimensional sentiment vectors of the impression of the article, presented

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I
SENTIMENT CATEGORY.

Sentiment Category	Word
Delight	Fun, Interesting, Good, Happy
Anger	Crazy, Irritating
Pain	Dowdy, Gammy, Gaunt, Sad
Fright	Terrible, Worried
Shame	Disgraceful, Squirmy
Like	Lovely, Honey, Pleasantly
Hate	Hateful, Gloomy
Frustration	Irritate, Vexation
Comfort	Relax, Simple, Sporty
Surprise	Abstracted, Amazement, Consternation

for information: "$happy$" ⇔ "$unhappy$", "$acceptance$" ⇔ "$rejection$", "$relaxation$" ⇔ "$strain$", "$anger$" ⇔ "$fear$".

As described above, many studies have extracted the emotion of a multi-dimensional vector. Using the system described herein, as the first step in responding to multi-dimensional vector, we specifically examine the P/N value ("$Positive$" ⇔ "$Negative$").

Turney et al.[7] calculate the sentiment orientations of words using the P/N value. Their P/N value is based on association between the word and the positive words, or between the word and negative words. Dave et al.[8] divide the reviews of web sites into positive reviews and negative reviews. Kobayashi et al.[9] propose a dictionary-based method for acquiring a P/N lexcon that specifies whether each entry means a positive or negative concept. Their method is based on a bootstrap method.

Fujimura et al.[10] classify a Web into positive or negative and extract a reputation from the Web. They propose a scoring method to extract reputation.

Takamura et al.[11] propose a method that calculates sentiment orientations of words. They regard sentiment orientations of words as spin of electron, and modeling the word network as a spin model. As described in this paper, we use their sentiment orientation of words. There are, however, many negative words in their sentiment orientations of words. We propose new method based on their sentiment orientations of words.

III. CALCULATING THE USER'S POSITIVE VALUE

A. Table of Semantic Orientation of Words

When we calculate a user's positive value, we use a "table of semantic orientation of words" proposed by Takamura[11]. The semantic orientation of words is a binary attribute: it is either positive or negative. For example "Good" and "Beautiful" are positive values, and "Bad" and "Dirty" are negative values. The semantic orientation of words is calculated using a lexicon network. In the semantic orientation of words, a negative word is "-1" and a positive word is "1". It consists of noun, verb, adjective, adverb, and auxiliary verb.

B. User's Positive Value

We calculate a user's positive value using current mood and desired mood that a user inputs. When we calculate a user's positive value, we use a table of semantic orientation of words. There are, however, 55,125 words in a table of semantic orientation of words. We use 60 words from it. Furthermore, we classify the 60 words into 10 categories using a sentiment dictionary[12] because a user selects a sentiment easily. Table 1 shows the 10 categories and some of the 60 words.

The user selects a current mood from 10 categories. Then the user selects a desired mood from Delight, Like, and Comfort because these three categories are positive categories. The user's positive value (PV) is the following.

$$PV = \frac{|S_2 - S_1|}{2} \quad (1)$$

Therein, S_1 denotes the current mood; S_1 is the desired mood. For example, when a user's current mood is "Sad(-0.9990)" and the desired mood is "Happy(0.998871)", then the PV is 0.9989505.

IV. WORD P/N VALUE

When we present a word-of-wisdom that fits a user's sentiment value, we quantify the word-of-wisdom. We designate it as the "word-of-wisdom P/N value". We calculate the word-of-wisdom P/N value using word P/N value for word-of-wisdom. When we calculate the word P/N value, first we use a table of semantic orientation of words. However, because many negative values are included in it, almost all word-of-wisdom P/N values become negative. We calculate the word P/N value using our method, and calculate the word-of-wisdom P/N value. Our method is based on the term frequency from each positive word-of-wisdom and negative word-of-wisdom.

A. Preliminary experiment1

We conducted a preliminary experiment to calculate the word P/N value. The flow of our preliminary experiment is the following:

1) We manually extracted 500 positive words-of-wisdom and 500 negative words-of-wisdom from the Internet.
2) There are eight subjects into which the 1000 words-of-wisdom were divided; they were further classified as positive or negative ones. We regard as positive or negative words-of-wisdom those for which five or more subjects have judged as positive or negative. Consequently, 377 words-of-wisdom were positive and 434 words-of-wisdom were negative.
3) We extracted 350 words-of-wisdom from those 377 positive words-of-wisdom and 434 negative words-of-wisdom. They become supervised data when we calculated the word P/N value.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE II
EXAMPLE OF EXTRACTING WORDS OF WISDOM IN OUR EXPERIMENT.

Positive Words of Wisdom
A man is not finished when he's defeated; he's finished when he quits.
While there's life, there's hope.
Our greatest glory consists not in never falling, but in rising every time we fall.
The more I learn, the more I realize I do not know. The more I realize I do not know the more I want to learn.
If you can dream it, you can do it.
Negative Words of Wisdom
I have generally found that a man who is good at excuses is usually good at nothing else.
The tragedy of life is what dies inside a man while he lives.
To mourn a mischief that is past and gone. Is the next way to draw new mischief on.
The most pathetic person in the world is someone who has sight, but has no vision.
Pay no attention to what the critics say; there has never been a statue set up in honor of a critic.

TABLE III
EXAMPLES OF STOP-WORDS.

part of speech	stop words
noun	man, life, we
adjectives	not
verb	do, have, has, are, is say think

TABLE IV
EXAMPLES OF WORD P/N VALUES.

word	word P/Nvalue
success	0.75
happy	0.723
change	0.26
laugh	0.25
loneliness	-0.51
unfortunate	-0.85
lost	-0.35
die	-0.80

Table 2 shows partial results obtained from our preliminary experiment.

B. Calculating Word P/N Value

We calculate a word P/N value according to the following steps:

1) We calculate each term frequency of nouns, adjectives, adverbs, and verbs in the results of 700 word-of-wisdom that we obtained in our preliminary experiment
2) We delete stop words for the results of step (1) because some words are not meant to be judged for P/N, such as nouns, verbs, and adjectives.
3) We calculate all words of words P/N WPN_i using the following equation.

$$WPN_i = \frac{X_{\mathrm{Pi}}}{X_{\mathrm{Pmax}}} - \frac{X_{\mathrm{Ni}}}{X_{\mathrm{Nmax}}} \tag{2}$$

Therein, X_{Pi} denotes the term frequency of word i in a positive words-of-wisdom, X_{Pmax} denotes a term frequency of word $Pmax$ which is the same part of speech as i and the maximum term frequency in a positive word-of-wisdom. Furthermore, X_{Ni} represents the term frequency of word i in a negative word-of-wisdom, $X_{\mathrm{Nmax}}$is a term frequency of word $Pmax$, which is the same part of speech as i and the maximum term frequency in negative word-of-wisdom. We calculate the word P/N value in advance; then we save the word in the database. We calculate and save 1933 words. Table 4 shows some word P/N values.

V. WORD-OF-WISDOM P/N VALUE

A word-of-wisdom consists of one sentence or multiple sentences. We propose a word-of-wisdom based on one sentence or multiple sentences.

A. Consisting of one sentence

We propose a method using only our proposed word P/N value. We calculate the word-of-wisdom P/N value MPN_j using the following formula.

$$MPN_j = \frac{\sum_{k=1}^{n} WPN_k}{n} \tag{3}$$

Therein, n represents number of words that have a word P/N value in word-of-wisdom j. Also, WPN_k is the word P/N value of the word.

B. Words-of-wisdom consisting of multiple sentences

Many words-of-wisdom comprise multiple sentences. In this case, when the word-of-wisdom is positive, not all sentences are positive. For example, Einstein says "I never think of the future. It comes soon enough." In this word-of-wisdom, the first sentence is negative sentence, but the second sentence is a positive sentence. Therefore, this word-of-wisdom is a positive word-of-wisdom. When the system ignores sentence structure, the system might judge the word-of-wisdom as negative. Then we analyze words-of-wisdom that comprise multiple sentences. We propose a formula of method of word-of-wisdom P/N values.

C. preliminary experiment2

We conducted a preliminary experiment to analysis to determine which sentences most influence the word-of-wisdom P/N value.

The experiment flow is as described below.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE V
RESULTS OF QUOTE ANALYSIS.

	Results matching word-of-wisdom result
Last sentence	88 %
Middle sentence	63 %
First sentence	63 %

TABLE VI
RESULT OF EXPERIMENT OF THE P/N VALUE.

	Positive word-of-wisdom		Negative word-of-wisdom	
	Recall	Precision	Recall	Precision
A	71%	54%	62%	52%
B	74%	55%	60%	51%

1) We randomly extracted 700 words-of-wisdom from the database. Of the 700 words-of-wisdom, 252 words-of-wisdom are simple sentences, 292 words-of-wisdom are two sentences, and 156 words-of-wisdom are three sentences.
2) The six subjects judged all sentences in 448 words-of-wisdom (292+156) as positive or negative according to their own viewpoint.
3) They also judged the 700 words-of-wisdom as positive or negative by their own viewpoint.

We analyze the results to determine the number of sentences that are the same as the results obtained using the words-of-wisdom system. Table V presents the results. Results show that 88% of the final sentence are the same as the P/N word-of-wisdom.

D. Word-of-wisdom P/N value of multiple sentences

We propose a formula of word-of-wisdom P/N value of multiple sentences using results of our preliminary experiment 2.

$$MPN_{fj} = \frac{\sum_{k=1}^{n-1} MPN_k + \alpha * MPN_{jn}}{n} \qquad (4)$$

In that equation, MPN_{fj} signifies the word-of-wisdom P/N value of multiple sentences word-of-wisdom fj. In addition, n denotes the number of sentence in fj, MPN_k represents the word-of-wisdom P/N value of each sentence in fj. MPN_{jn} stands for the last sentence of word-of-wisdom P/N value. α denotes the weight of the last sentence. In our preliminary experiment 2, the last sentence effect the word-of-wisdom is 88%; the other sentence effect the word-of-wisdom is 63%. That is, the last sentence has a 1.4 times greater effect on the result than other sentences do. Therefore, we regard α as 1.4.

VI. PROTOTYPE SYSTEM

We developed a prototype system using Ruby on Rails and implemented the prototype.
The flow of the system is explained below and is shown in Figure1.

1) A user selects an interesting category from 10 categories in our prototype system: “Sports”, “Manga”, “Movie”, “TV-Drama”, “Love”, “Friends”, “Effort”, “Adolescence”, “Life”, and “Politics”.
2) The user selects a current mood from 10 semantic categories.
3) The user selects a desired mood from three categories.
4) The system calculates a user’s positive value based on the current mood and desired mood.
5) The system extracts a word-of-wisdom that has the most similar value to a user’s positive value from the word-of-wisdom database.

The system calculates the word P/N value and word-of-wisdom P/N value and create a word-of-wisdom database in advance.

VII. EXPERIMENTS

We conducted two experiments. One experiment was conducted to examine the extraction of the word-of-wisdom P/N value. We next tested the usability of our proposed system.

A. Experiment 1: Extraction of word-of-wisdom P/N value

In the experiment, our purpose is measuring the benefits of our proposed method. We used 200 words-of-wisdom taken from the Internet randomly. We classified the words-of-wisdom manually as positive or negative. We regard those as correct results. Subsequently, the system classified the words-of-wisdom as positive or negative, and calculated recall and precision. The system calculated them using two methods. One is using only the word P/N value (method A), the other is using the word P/N value (method B) and handling method for multiple sentences.

Table 6 presents results of our experiments. The results of recall are good; multiple sentence cases are handled especially better than those obtained when using only the word P/N value. Results show that our proposed multiple sentence method is beneficial. The results of precision, however, are not good. The reason is that our method does not address contradiction of the sentences. For example, “I never think of the future. It comes soon enough.’ in a first sentence is contradiction sentence, but we do not care about it. In our system, this case becomes a positive sentence. We should consider contradictory sentences.

B. Experiment 2: User experiment

We performed user experiments involving our system. There were five subjects. We prepare situations of three types: an ordinary situation, positive situation, and negative situation. Then subjects answered the questions in the situation. The flow of experiments is as follows:

1) Subjects use our system and answer the question.
2) They talk about a cheerful theme.
3) Subjects use our system and answer the question again.
4) They talk about an unhappy theme.
5) Subjects use our system and answer the question again.

The questions are the following:

Q1 How did your feeling change after using our system?
Q2 Did the system presented the word-of-wisdom that you wanted it to present?

Figure 2 presents results of Q1. Figure 3 shows the results of Q2. In Figure 3, result 1 is shown as very bad. In fact, result 5 is the best match for the word-of-wisdom. The results of Q1 show good results in each situation. Before

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

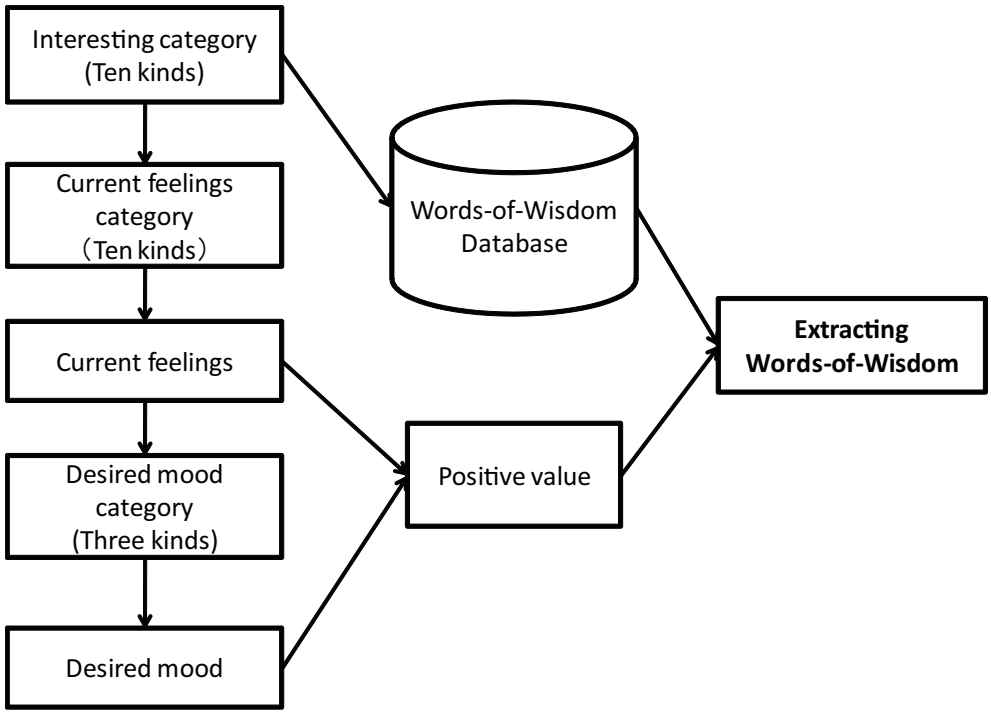

Fig. 1. System Flow.

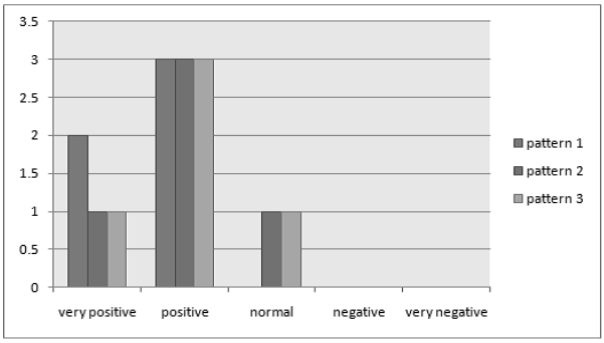

Fig. 2. Result of Q1.

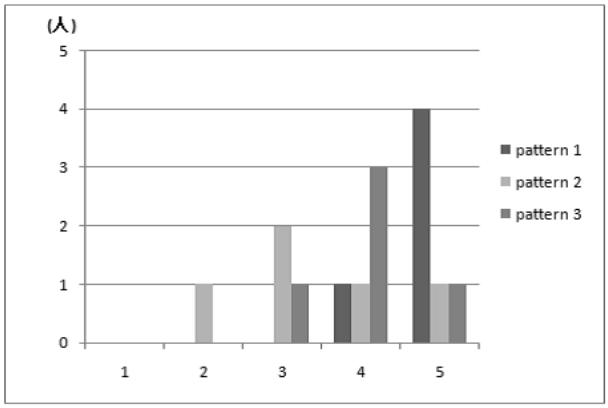

Fig. 3. Result of Q2.

experiments, we estimate that our system is best in a negative situation. However, the results of Q1 shows good results in each situation, especially in ordinary situations. In Q2, results of the ordinary situation and the negative situation are good. However, the result of positive situation is not as good.

VIII. Conclusion

As described in this paper, we proposed a system that presents word-of-wisdom. Our system presents word-of-wisdom that matches a user's desired sentiments. We also proposed a method of calculating the word P/N value for words-of-wisdom and the word-of-wisdom P/N value based on it. Then, we created a word-of-wisdom database based on our word-of-wisdom P/N value. Using our proposed system, first a user inputs a current mood and a desired mood. Then the system calculates user's P/N value. Then the system extracts a word-of-wisdom that has a similar P/N value to that of the user from the word-of-wisdom database. Subjects for future work include the following.

- Improved recall ratio in the word-of-wisdom P/N value
  As described in this paper, we ignore negative words such as "not" or "never". In the future, we must consider such negative words to calculate the sentiment of a word-of-wisdom.
- Consideration of other vectors
  As described in this paper, we propose the PN vector for word-of-wisdom as a first step of extracting sentiment from a word-of-wisdom. In the near future, we plan to extract words-of-wisdom using other multiple sentiment vectors.

References

[1] V.Hatzivassiloglou, K.R.McKeown, "Predicting the semantic orientation of Adjectives," In Proceedings of the Thirty-Fifth Annual Meeting of the European Chapter of the Association for Computational Linguistics for Computational Linguistics", pp.174-181, 1997.
[2] J.Leskovec, D.Huttenlocher, J.Kleinberg, "Predicting Positive and Negative Links in Online Social Networks," WWW 2010, pp.641-650, April 26–30, 2010

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[3] R.Plutchik, “The Multi factor-Analytic Theory of Emotion,” Psychology, Vol.50, pp.153-171, 1960.
[4] R.Plutchik, “The nature of emotions,” American Scientist, Vol.89, pp.344–350, 2001.
[5] T.Kumamoto, K.Tanaka, “Proposal of Impression Mining from News Articles,” Lecture Notes in Computer Science, Vol.3681, pp.901-910, 2005.
[6] Y.Kawai, T.Kumamoto, K.Tanaka, “Fair News Reader: Recommending News Articles with Different Sentiments Based on User Preference,” Lecture Notes in Computer Science, LNAI 4692, pp.612-622, 2007.
[7] P.D.Turney, “Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews,” Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), pp.417-424, 2002.
[8] K.Dave, S.Lawrence, D.M.Pennock, “Mining the peanut gallery: opinion extraction and semantic classification of product reviews,” In Proceedings of the 12th International World Wide Web Conference, 2003.
[9] N.Kobayashi, T.Inui, K.Inui, “Dictionary-Based Acquisition of the Lexical Knowledge for p/n Analysis (in Japanese),” In Proceedings of the Japanese Society for Artificial Intelligence, SLUD-33, pp.45-40, 2001.
[10] S.Fujimura, M.Toyoda, M.Kitsuregawa, “A Reputation Extracting Method Considering Structure of Sentence (in Japanese),” DEWS2005, 6C-i8, 2005.
[11] H.Takamura, T.Inui, M.Okumura, “Extracting semantic orientations of words using spin model,” In Proceedings 43rd Annual Meeting of the Association for Computational Linguistics (ACL’05), pp.133-140, 2005.
[12] A.Nakamura, “Sentiment Dictionary,” Tokyodo Publisher, 1993.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)