WellnessWordNet : A Word Net for Unconstrained Subjective Well-Being Monitor ing Based on Unstructured Data and Contextual Polar ity

웰니스워드넷 : 비정형데이터와 상황적 긍부정성에 기반하여 주관적 웰빙 상태를 무구속적으로 모니터링하기 위한 워드넷 개발

저자 (Authors): Yeongeun Song, Suhyun Nam, Ohbyung Kwon

출처 (Source): 지능정보연구 22(3), 2016.9, 1-21 (21 pages)
Journal of Intelligence and Information Systems 22(3), 2016.9, 1-21 (21 pages)

발행처 (Publisher): 한국지능정보시스템학회
Korea Intelligent Information Systems Society

URL: http://www.dbpia.co.kr/Article/NODE07016491

APA Style: Yeongeun Song, Suhyun Nam, Ohbyung Kwon (2016). WellnessWordNet : A Word Net for Unconstrained Subjective Well-Being Monitor ing Based on Unstructured Data and Contextual Polar ity. 지능정보연구, 22(3), 1-21.

ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2016 September: 22(3): 1~21 http://dx.doi.org/10.13088/jiis.2016.22.3.001

WellnessWordNet: A Word Net for Unconstrained Subjective Well-Being Monitoring Based on Unstructured Data and Contextual Polarity*

Yeongeun Song
School of Management, Kyung Hee University
(*dudeun13@khu.ac.kr*)

Suhyun Nam
School of Management, Kyung Hee University
(*gloomycloud@khu.ac.kr*)

Ohbyung Kwon
School of Management, Kyung Hee University
(*obkwon@khu.ac.kr*)

IT-based subjective well-being (SWB) services, a main part of wellness IT, should measure the SWB state of individuals in an unrestrained, cost-effective manner. The dictionaries for sentiment analysis available in the market may be useful for this purpose, but obtaining proper sentiment values using only words from the sentiment lexicon is impossible; therefore, a new dictionary including wellness vocabulary is needed. The existing sentiment dictionaries link only a single sentiment value to a single sentiment word, although sentiment values may vary depending on personal traits. In this study, we develop an extended version of the SenticNet sentiment dictionary dubbed WellnessWordNet. SenticNet is considered the best and most expressive among the already existing sentiment dictionaries. Using the information provided by SenticNet, we created a database including the wellness states (estimated values) of stress, depression, and anger to develop the WellnessWordNet system. The accuracy of the system was validated through actual tests with live subjects. This study is unique and unprecedented in that i) an extended sentiment dictionary, WellnessWordNet, is developed; ii) values for wellness state language are offered; and iii) different sentiment values, namely contextual polarity, for people of the same gender or age group are suggested.

Key Words : Subjective Well-Being; Sentiment Analysis; Contextual Polarity; Unstructured Data; SenticNet

Received : July 20, 2016 Revised : August 10, 2016 Accepted : August 11, 2016
Publication Type : Regular Paper Corresponding Author : Ohbyung Kwon

1. Introduction

In recent research, sentiment analysis has emerged as a key factor in evaluating corporations, products, and services, drawing increased attention from both industry practitioners and academics. In sentiment analysis, the words or word groups in a text are recognized as having either positive or negative affect. Several sentiment analysis dictionaries have been developed for this purpose.

* This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.R0126-15-1007, Curation commerce based global open market system development for personal happiness enhancement)

However, currently available lexical resources for opinion polarity and affect recognition, such as SentiWordNet (Esuli and Sebastiani, 2006) or WordNet-Affect, are known to be rather noisy and limited. These resources include only dictionary words; they do not reflect emotional (everyday) language that relates to common sense knowledge.

SenticNet (Cambria et al., 2012) was developed to overcome this drawback. SenticNet is a lexical resource that was built by clustering a vector space of affective common sense knowledge. It lists several thousand concepts along with their polarity. SenticNet concepts are composed of a concept string, five semantics, four sentics (i.e., Pleasantness, Attention, Sensitivity, and Aptitude), and polarity (Xia et al., 2014).

However, for the purposes of measuring SWB (subjective well-being), SenticNet needs to be complemented in some respects. First, the four sentics adopted by SenticNet indicate emotional state only, making it difficult to measure other mental health concepts such as depression, stress, or fatigue. Emotional state is closely related to mental health, but they are not identical. Second, like other sentiment analysis dictionaries, SenticNet does not reflect individual personal traits. According to the findings of conventional studies, mental health state is known to be associated with gender (Shmotkin, 1990), income (McBride, 2001), number of friends (Diener, 2000), religious faith (Diener, 2000; Dolan et al., 2008), and age (Smith et al., 2002). Therefore, polarity that does not consider personal traits or individual situations cannot be accurate enough.

In this study, we develop an extended version of the SenticNet sentiment dictionary dubbed WellnessWordNet. SenticNet is considered the best and most expressive among the various sentiment analysis dictionaries. Using the information provided by SenticNet, the wellness states (estimated values) of stress, depression, fatigue, and anger were added to develop WellnessWordNet. Among various possible personal traits, age and gender were included as characteristics related to SWB state. The accuracy of the developed system was validated through actual tests with live subjects.

This study consists of five sections: section 2 introduces existing sentiment analysis dictionaries and the concept of subjective well-being, and section 3 describes how we built WellnessWordNet. The performance of WellnessWordNet is compared with other methods in section 4, and in the last section, the significance, limitations, and conclusions of this study are discussed.

2. Related Studies

2.1 Sentiment Analysis Dictionary

The sentiment dictionary is central to semantic analysis for the purposes of (text-based) text classification, sentiment analysis, and opinion mining, in addition to lexical analysis (e.g., part-of-speech tag analysis, pre-processing). Currently available sentiment dictionaries are listed in Table 1. Three strategies can be utilized to build a sentiment dictionary: i) adding to WordNet domains through extensions such as WordNet-Affect and SentiWordNet; ii) establishing an ontological model to represent emotional concepts by type and intensity (e.g., EmotionML and Onyx); and iii) building polarity of emotions based on people's everyday, common sense language (e.g., SenticNet). Among the available tools, SenticNet is currently one of the most comprehensive, freely available semantic resources for opinion mining. Compared to its peers, SenticNet covers the multidimensionality of emotions, enabling more precise sentiment analysis. SenticNet is outstanding in that it includes results not only for a single word, but for multiple words.

As most sentiment dictionaries other than SenticNet focus on emotions only, they do not provide polarity values for concepts based on emotions. In addition, people's state of wellness or quality of life is definitely connected to their emotions, but results of analyses based on these dictionaries do not reflect this fact. These dictionaries suggest unified values for unspecified individuals. Depending on various factors such as age or gender, however, a single word may be used for different sentiment values. Because they provide only one value per word, the accuracy of analyses using these dictionaries is limited.

2.2 Subjective Well-being (SWB)

Subjective well-being (SWB) is a construct used to evaluate quality of life. It includes cognitive judgment and affective judgment, and is basically synonymous with life satisfaction, happiness, positive and negative emotionality, and psychological well-being. Available tools for measuring SWB include the Life Satisfaction

〈Table 1〉 Sentiment dictionaries

Dictionary	Description
SentiWordNet	An enhanced lexical resource for sentiment analysis and opinion mining related to WordNet (Hung and Lin, 2013).
SenticNet	A lexical resource built by clustering a vector space of affective common sense knowledge, which lists several thousand concepts along with their polarity (Cambria et al., 2012).
WordNet-Affect	WordNet-Affect is an extension of WordNet Domains, including a subset of synsets suitable to represent affective concepts correlated with affective words (Strapparava and Valitutti, 2004).
HowNet	Calculates the similarity between words and the semantic relevance between words (Liu & Li, 2002).
Affective Norms for English Words (ANEW)	Provides a set of normative emotional ratings for a large number of words in English and a set of verbal materials that have been rated in terms of pleasure, arousal, and dominance and manually annotated (Bradley and Lang, 1999).
EmotionML	Offers twelve vocabularies for categories, appraisals, dimensions, and action tendencies. A vocabulary is a set of possible values for any given attribute of the emotion (Ashimuraetal et al., 2012). Users can either define their own vocabularies or reuse one of the existing ones.
Onyx	A semantic vocabulary of emotions with a focus on lexical resources and emotion analysis services (Sánchez-Rada and Iglesias, 2016).

Index-Z (Kozma and Stones, 1987), Life Satisfaction Index-A, the Philadelphia Geriatric Center Morale Scale (PGCMS) (Lawton, 1975), Satisfaction With Life Scale (SWLS) (Diener et al., 1985) Memorial University of Newfoundland Scale of Happiness (MUNSH) (Zhang et al., 2016), General Well-Being (GWB) scale (Dupuy, 1978), Life Satisfaction for Elderly Scale (Pezzuti et al., 2015), and Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988), all of which are based on the measures of Havighurst and Neugarten (Havighurst et al., 1961). All of the methods above, however, rely on questionnaires in which respondents must answer multiple questions or complete a self-report. It is impossible to apply such methods to SWB measurement; rather, SWB should be measured in a repeated and swift manner online. Accordingly, it is appropriate to consider a sentiment analysis dictionary as a tool to measure SWB automatically using unstructured data. This is an unrestrained, cost-effective way of measuring SWB services that minimizes manual data input by users and its accompanying errors.

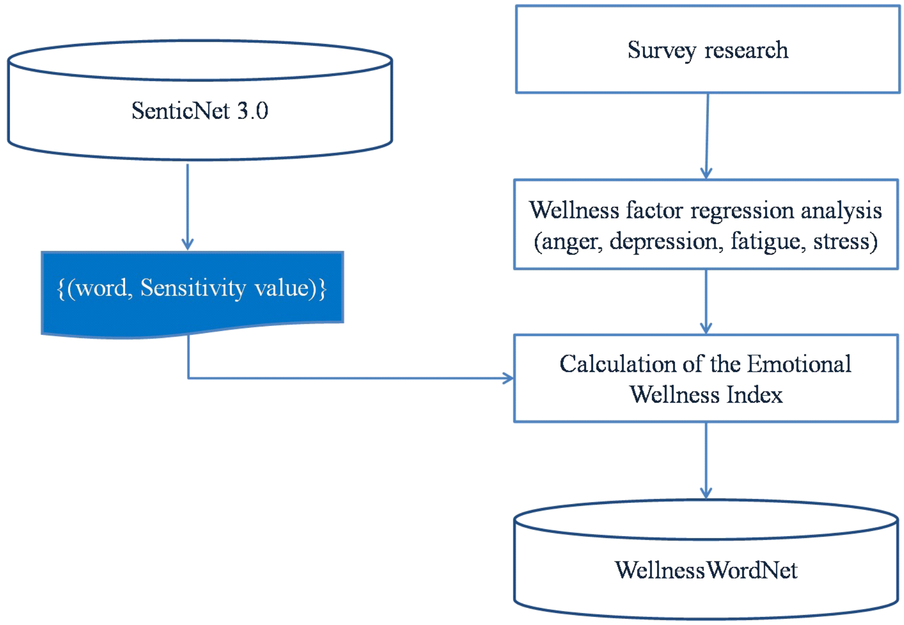
3. Construction of WellnessWordNet

3.1 Overall Process

There are four reasons why SenticNet was adopted for building WellnessWordNet. First, unlike its counterparts that offer only positive or negative values, SenticNet delivers multidimensional sentiment values. The psychological emotional state is classified into four dimensions, and each emotion category is divided based on the hourglass model. SenticNet contains the emotion of Sensitivity, while its counterparts do not. By setting Sensitivity as an independent variable and other wellness state values as dependent variables, regression analysis can be used to estimate an individual's wellness state. Second, SenticNet utilizes a corpus-based method that combines both a dictionary and common sense knowledge. With this method, any ambiguity issues arising in certain domains are resolved.

The system architecture for building WellnessWordNet is illustrated in Figure 1. First, the four sentiment values (per type) of each sentiment lexicon registered in SenticNet were identified. After checking the sentiment value of "Sensitivity (wellness state)" and the relevant vocabulary, a set { (word, Sensitivity value) } was created.

We conducted a survey using a questionnaire in which four wellness states (anger, depression, fatigue, and stress) together with the respondents' profiles (age, gender) were included. Questions frequently used in positive psychology were also used in our survey. Based on the results of the survey, the degree of anger and the respondents' traits were set as independent variables, and the rest states (depression, fatigue, and stress) became dependent variables in the regression analysis. Through the analysis, statistically significant regression equations were derived. With these equations in place, we then substituted Sensitivity values in the set of { (word, Sensitivity value) }



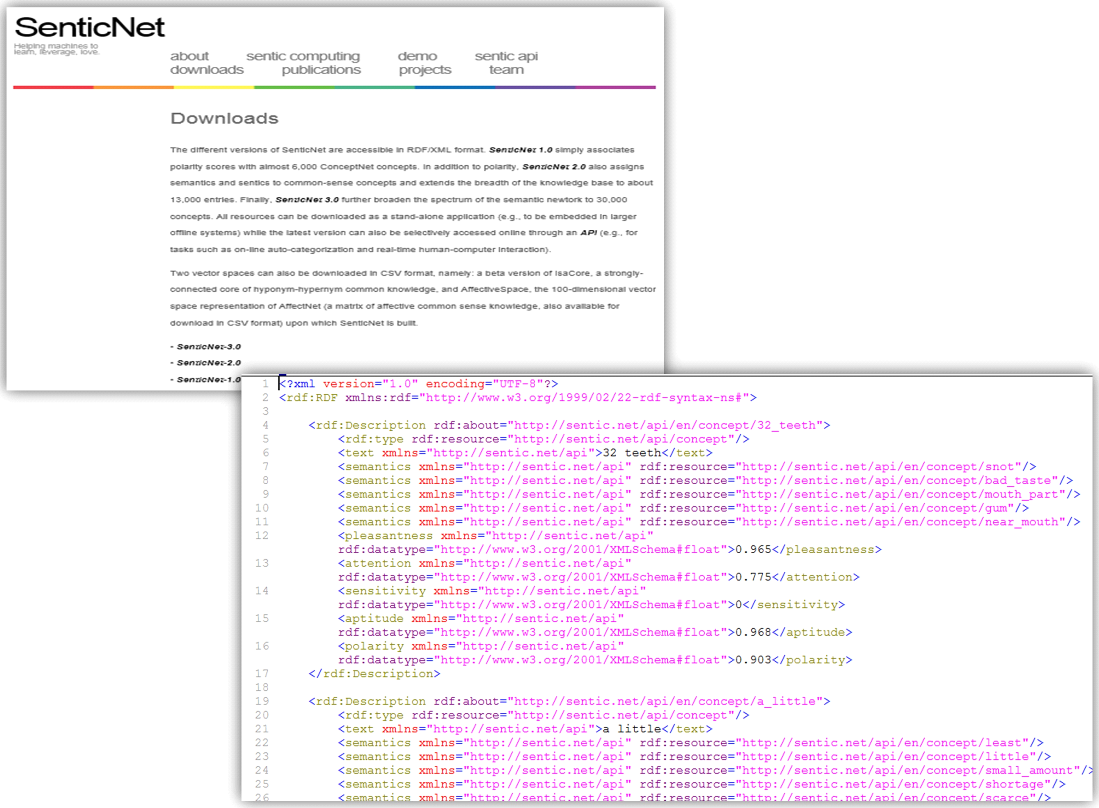
〈Figure 1〉 System architecture for building WellnessWordNet

with the independent variables. Thus, the estimated values of depression, fatigue, and stress for the corresponding word were obtained. Three values can be earned: i) when Sensitivity value only is known; ii) when Sensitivity value and gender are known; and ii) when Sensitivity value and age are known. In the last stage, we finalized and created WellnessWordNet, an extended version of SenticNet, including newly-obtained estimated values.

3.2 SenticNet Information Collection

First, we downloaded the sentiment lexicons of SenticNet v. 3.0 (refer to Figure 2). It has about 30,000 sentiment lexicons built in XML format, including multi-word lexicons (Cambria and Hussain, 2012)

The affective dimensions of the collected SenticNet data, namely Pleasantness, Attention, Sensitivity, and Aptitude, are divided by the hourglass model and listed in Table 2. Each affective dimension is characterized by six levels of activation (measuring the strength of an emotion), termed “sentic levels,” which represent the intensity thresholds of the expressed/perceived emotion (Cambria and Hussain, 2012; Cambria et al., 2012; Cambria et al., 2011; Cambria., 2016). Sensitivity’s sentic levels are: rage, anger, annoyance, apprehension, fear, and terror. Therefore, Sensitivity has information on anger. For this study, we therefore extracted vocabularies related to Sensitivity.



〈Figure 2〉 Collection of SenticNet data

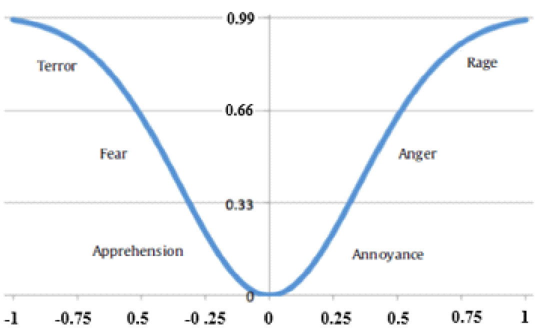
〈Table 2〉 The hourglass model

Interval	Pleasantness	Attention	Sensitivity	Aptitude
[G(1), G(2/3))	ecstasy	vigilance	rage	admiration
[G(2/3), G(1/3))	joy	anticipation	anger	trust
[G(1/3), G(0))	serenity	interest	annoyance	acceptance
(G(0), –G(1/3)]	pensiveness	distraction	apprehension	boredom
(–G(1/3), –G(2/3)]	sadness	surprise	fear	disgust
(–G(2/3), –G(1)]	grief	amazement	terror	loathing

Sentic levels are based on the Gaussian function $G(x) = -\frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$ and arranged in a symmetric inverted bell curve shape. The sentiment values range between 1 and -1. The sensitivity emotional flow is illustrated in Figure 3.

〈Table 3〉 Sample scores of some Sensitivity words in SenticNet

Interval	Number of words	Sample words
-1 ~ -0.661	584	awe, coercion, pass out, reckless, terrible, revenge, scare, sickness, etc…
-0.66 ~ -0.331	323	regretful, urgency, dysphoric, self-reproach, feel guilt, remorseful, etc…
-0.33 ~ -0.01	6026	tense, fearless, quietness, encouraged, ashamed, deplorable, unhealthy, calmly, etc…
0	11,631	absorbing, accessible, adventurer, backward, embassy, emergence, literacy, etc…
0.01 ~ 0.33	4318	feel embarrassed, penniless, contradiction, stabilize, waste of time, feel exhausted, melodic noise, etc…
0.331 ~ 0.66	2114	indisposition, aggressiveness, fatigue, parsimonious, hideousness, irksome, fuggy, etc…
0.661 ~ 1	5004	rage, destroy evidence, insomnia, retaliatory, infelicitous, etc…



〈Figure 3〉 Sensitivity emotional flow

The Sensitivity information obtained at a designated interval of SenticNet (extracted) is provided in Table 3. In total, 11,436 words with Sensitivity > 0 were obtained, most of which expressed intense anger, and 6,933 words (Sensibility < 0) represented defensive attitudes related to anger.

3.3 Regression Analysis for Derivation of Wellness Values

3.3.1 Data Collection

Data was collected by an online survey agency for 6 days (Mar. 18–23, 2016) among 574 men and women in their 20s to 50s. The questionnaire included questions related to stress, depression, anger, and fatigue, the four factors required to measure the demographic characteristics and mental–emotional health in wellness. The Likert 7-point scale (1: strongly disagree, 7: strongly agree) was applied. Table 4 presents the demographic characteristics of the respondents.

3.3.2 Data Validation

After collecting data for the two measures of depression, we utilized the Center for Epidemiological Studies-Depression Scale (CES-D) and Hamilton Rating Scale for Depression (HRSD) to analyze the data. The Korea Depression Scale (KDS), a complementary version of the CES-D and HRSD translated by Lee and Rhee(2003), was also used to make up a total of 15 questions. In the case of anger, the data was measured using the State-Trait Anger Expression Inventory (STAXI), which was created by Spielberger (1988), and from its Korean version, STAXI-K, 8 questions on trait anger were extracted to be part of the

〈Table 4〉 Demographic characteristics of subjects

Characteristics	Categories	Participants	%
Gender	Male	288	50.2
	Female	286	49.8
Age group	20–29	236	41.1
	30–39	178	31.0
	40–49	105	18.3
	Over 50	55	9.6
Occupation	Employed	343	59.8
	Unemployed	231	40.2
Area	Outside capital area	194	33.8
	Within capital area	380	66.2
Area type	Rural	29	5.1
	Urban	545	94.9
Hobbies	Yes	306	53.3
	No	268	46.7
Number of family members	1	67	11.7
	2	92	16.0
	3	168	29.3
	4	197	34.3
	> 5	50	8.7

questionnaire for this study. For fatigue, the Fatigue Severity Scale (FSS) developed by Krupp et al. (1989) was used to generate 7 questions. Lastly, the A/B Lifestyle questionnaire was adopted for measurement of stress (Charlesworth et al., 1984).

For data validation, the SPSS 20.0 program was utilized to conduct an exploratory factor analysis and reliability analysis. Table 5 presents the results of the analyses. For the exploratory factor analysis, major component factors were analyzed according to the varimax method (rotation), and factors with eigenvalues ≥1.0 were extracted. As the testing of the fit of the EFA model resulted in KMO = 0.944 and Bartlette's test = 20165.386 (p = 0.000), the post hoc multiple comparisons were considered to be no problem. Reliability values for each factor were as follows: depression = 0.968, anger = 0.942, fatigue = 0.904, stress = 0.882. Therefore, all factors demonstrated sufficient reliability.

3.3.3 Wellness Values Estimated by Regression Analysis

By conducting a regression analysis based on the survey data, we obtained statistically meaningful regression equations. Table 6 shows the results after including the profile data (i.e., gender, age) of respondents for testing the relation

〈Table 5〉 Results of factor analysis

Factor name	ITEM	factor loading			
		1	2	3	4
Depression	My life is regrettable and distressed	.868	.168	.057	.026
	I feel like I'm a loser	.864	.164	.086	-.014
	My life is all vanity and meaningless	.848	.186	.178	.103
	I have crying spells or feel like crying	.846	.197	.085	-.045
	I think I have no value and am ashamed of myself	.831	.236	.112	-.047
	I think my life has been a failure	.822	.219	.033	-.017
	I spend most of my time feeling blue	.818	.233	.177	.125
	I feel depressed	.803	.131	.193	.092
	I feel helpless a lot of the time	.800	.165	.308	.018
	I feel that I could not shake off the blues even with help from my family or friends	.770	.199	.125	.042
	Recently, I have lost motivation to solve problems	.768	.162	.293	.067
	I feel that everything I do is an effort	.726	.134	.325	.103
	I have frequent crying spells	.698	.155	.190	.075
	I have trouble keeping my mind on what I am doing	.683	.175	.340	.059
	I talk less than usual	.673	.135	.277	.106
Anger	I can't control my anger	.214	.864	.093	.107
	I am a hotheaded person	.214	.859	.107	.055
	I have a fiery temper	.212	.858	.098	.042
	I lose my temper easily	.264	.855	.116	.029
	I get angry very quickly	.178	.831	.107	.052
	I am easily angered	.207	.816	.105	.178
	It makes me furious when I am criticized in front of others	.217	.682	.068	.083
	I get angry when I am slowed down by others' mistakes	.220	.651	.081	.076
Fatigue	Fatigue causes frequent problems for me	.216	.088	.840	.094
	Fatigue interferes with my work, family, or social life	.227	.125	.835	.081
	My fatigue prevents sustained physical functioning	.222	.109	.833	.061
	Fatigue interferes with carrying out certain duties and responsibilities	.251	.135	.831	.028
	Fatigue interferes with my physical functioning	.260	.108	.764	.075
	I am easily fatigued	.295	.079	.635	.103
	I bring work home or work late	.080	.055	.545	.208
Stress	I feel that I must get things finished	-.037	-.032	.035	.807
	I strive to do a good job	.168	.077	.186	.793
	I think often about what to do next	.007	.011	.067	.760
	I always feel responsible	-.087	-.034	.067	.719
	I often finish sentences for others	-.038	.038	.042	.716
	I pay careful attention to detail	.037	-.024	-.045	.711
	I am frequently hurried for appointments	.192	.218	.136	.643
	I am always in a hurry	.168	.265	.100	.617
	I want to be recognized by others	.052	.171	.114	.584
Eigenvalue		10.229	5.914	4.899	4.749
% of variance		26.227	15.165	12.562	12.176
Cumulative %		26.227	41.392	53.954	66.130
Cronbach's α		0.968	0.942	0.904	0.882
KMO = 0.944, Bartlette's test (χ^2 = 20165.386, df = 741, p-value = 0.000)					

〈Table 6〉 Results from regression analysis: depression

Variable	Model Ⅰ		Model Ⅱ		Model Ⅲ		Model Ⅳ	
	β	t	β	t	β	t	β	t
Anger	.872***	62.955	.845***	46.036	.880***	60.69	.852***	45.678
Gender			.221**	2.202	-	-	.234***	2.333
Age					-.327*	-1.807	-.355**	-1.964
F	3963.359		1997.422		1991.142		1339.558	
R^2	0.874		0.875		0.874		0.876	
Adjusted R^2	0.873		0.874		0.874		0.875	

* p < 0.1, ** p < 0.05, *** p < 0.01

between demographic factors and anger and depression. In model I, only anger was included; in model II, anger and gender were included; in model III, anger and age were assigned as independent variables; and in model IV, all independent variables from models I–III were included.

In the regression analysis with depression as the dependent variable, it was found that: in model I, anger had a statistically significant effect on depression (β = .872), and the total explanatory power of this variable was about 87.4%; in model II, anger (β = .845) and gender (β = .221) had a significant positive effect on depression, and their total explanatory power was 87.5%; in model III, anger (β = .880) had a positive effect, while age (β = -.327) had a negative effect, with a total explanatory power of about 87.4%; in model IV, with all independent variables added, anger (β = .852) and gender (β = .234) had a positive effect, while age (β = -.355) had a negative effect on depression, with a total explanatory power of 87.6%.

Table 7 demonstrates the results of the analysis after including the profile data (i.e., gender, age) of respondents for testing of the relation between demographic factors and anger and fatigue. In model I, only anger was included; in model II, anger and gender were included; in model III, anger and age were assigned as independent variables; and in model IV, all independent variables from models I–III were included.

〈Table 7〉 Results from regression analysis: fatigue

variable	Model Ⅰ		Model Ⅱ		Model Ⅲ		Model Ⅳ	
	β	t	β	t	β	t	β	t
anger	1.044***	64.189	0.964***	45.794	1.055***	62.114	0.975***	45.73
gender			0.663***	5.77	-	-	0.684***	5.971
age					-0.484**	-2.278	-0.566***	-2.737
F	4120.223		2192.846		2077.773		1480.99	
R^2	0.878		0.885		0.879		0.886	
Adjusted R^2	0.878		0.884		0.879		0.886	

* p < 0.1, ** p < 0.05, *** p < 0.01

After the regression analysis with fatigue as the dependent variable, it was found that: in model I, anger had a statistically significant effect on fatigue (β = 1.044), with a total explanatory power of about 87.8%; in model II, anger (β = .964) and gender (β = .663) had a significant and positive effect on fatigue, with a total explanatory power of 88.5%; in model III, anger (β = 1.055) had a positive effect, while age (β = -.484) had a negative effect, with a total explanatory power of about 87.9%; lastly, in model IV with all independent variables included, anger (β = .975) and age (β = .684) had a positive effect, while age (β = -.566) had a negative effect on fatigue, with a total explanatory power of 88.6%.

Table 8 provides the results of the analysis after including the profile data (i.e., gender, age) of respondents for testing of the relation between demographic factors and anger and stress. In model I, only anger was included; in model II, anger and gender information were included; in model III, anger and age were assigned as independent variables; and in model IV, all independent variables from models I–III were included.

The results of the regression analysis with stress as the dependent variable revealed that: in model I, anger had a statistically significant effect on stress (β = 1.206), with a total explanatory power of 89.1%; in model II, anger (β = 1.096) and gender (β = 0.908) had a significant positive effect on stress, with a total explanatory power of 90.0%; in model III, anger (β = 1.128) and age (β = -.624) had a significant effect; in model IV with all independent variables included, anger (β = 1.047), gender (β = .829), and age (β = -.472) had a positive effect on stress, with a total explanatory power of 90.0%.

3.3.4 Output Results of Wellness Polarity Testing

Regression equations (1) to (15) were derived from the regression analysis:

$$\text{Depression} = 0.872 * \text{Sensitivity} \quad (1)$$

$$\text{Fatigue} = 1.044 * \text{Sensitivity} \quad (2)$$

$$\text{Stress} = 1.206 * \text{Sensitivity} \quad (3)$$

$$\text{depression gender (female)} = 0.845 * \text{Sensitivity} \quad (4)$$

$$\text{depression gender (male)} = 0.845 * \text{Sensitivity} + 0.221 \quad (5)$$

〈Table 8〉 Results from regression analysis: stress

variable	Model Ⅰ		Model Ⅱ		Model Ⅲ		Model Ⅳ	
	β	t	β	t	β	t	β	t
anger	1.206***	68.355	1.096***	48.866	1.128***	47.254	1.047***	40.34
gender			0.908***	7.413	-	-	0.829***	6.729
age					0.624***	4.749	0.472***	3.665
F	4972.438		2583.638		2435.357		1719.62	
R^2	0.891		0.900		0.895		0.900	
Adjusted R^2	0.891		0.900		0.895		0.900	

* p < 0.1, ** p < 0.05, *** p < 0.01

$$\text{depression\_age } (< 50) = 0.88 * \text{Sensitivity} \tag{6}$$

$$\text{depression\_age } (\geq 50) = 0.88 * \text{Sensitivity} - 0.327 \tag{7}$$

$$\text{fatigue\_gender (female)} = 0.964 * \text{Sensitivity} \tag{8}$$

$$\text{fatigue\_gender (male)} = 0.964 * \text{Sensitivity} + 0.663 \tag{9}$$

$$\text{fatigue\_age } (< 50) = 1.055 * \text{Sensitivity} \tag{10}$$

$$\text{fatigue\_age } (\geq 50) = 1.055 * \text{Sensitivity} - 0.484 \tag{11}$$

$$\text{stress\_gender (female)} = 1.096 * \text{Sensitivity} \tag{12}$$

$$\text{stress\_gender (male)} = 1.096 * \text{Sensitivity} + 0.908 \tag{13}$$

$$\text{stress\_age } (< 30) = 1.128 * \text{Sensitivity} \tag{14}$$

$$\text{stress\_age } (\geq 30) = 1.128 * \text{Sensitivity} + 0.624 \tag{15}$$

Formulas (1)–(3) are based on simple regression analyses expressing only values for stress, fatigue, and depression. Formulas (4)–(7) were derived from multiple regression analyses of depression using personal traits such as gender and age; men were vulnerable to depression more than women by 0.221, and with age included, men and women in their 50s or older were vulnerable to depression at -0.327. The same was true for formulas (8)–(11) with fatigue; men were more vulnerable than women to fatigue by 0.663, and men and women in their 50s or older even more so at -0.484. Formulas (12)–(15) resulted from multiple regression analyses of stress; again, a greater effect was observed for men than for women for the stress variable by 0.908, and men and women in their 30s had higher scores for stress at 0.624 than respondents in other age groups.

Using these regression equations and the set of { (word, Sensitivity value) } extracted from SenticNet, Sensitivity values were assigned to independent variables to obtain estimated values for depression, fatigue, and stress. Figure 4 demonstrates the estimated wellness outputs.

By converting the derived estimates into an XML format file and printing it out, WellnessWordNet was generated. An example of the finalized WellnessWordNet is provided in Appendix A.

WellnessWordNet	Sensitivity	depression	fatigue	stress	depression_gender		depression_ages		fatigue_gender		fatigue_ages		stress_gender		stress_ages	
					female	male	<50	>=50	female	male	<50	>=50	female	male	<30	>=30
a lot of fat	0.05	0.0436	0.0522	0.0603	0.04225	0.26325	0.044	-0.283	0.0482	0.7112	0.05275	-0.43125	0.0548	0.9628	0.0564	0.6841
a lot of flowers	0.025	0.0218	0.0261	0.03015	0.021125	0.242125	0.022	-0.305	0.0241	0.6871	0.026375	-0.45763	0.0274	0.9354	0.0282	0.65405
a lot of information	-0.05	-0.0436	-0.0522	-0.0603	-0.04225	0.17875	-0.044	-0.371	-0.0482	0.6148	-0.05275	-0.53675	-0.0548	0.8532	-0.0564	0.5639
a lot of noise	0.896	0.781312	0.935424	1.080576	0.75712	0.97812	0.78848	0.46148	0.863744	1.526744	0.94528	0.46128	0.982016	1.890016	1.010688	1.700992
a lot of stress	-0.104	-0.090688	-0.108576	-0.12542	-0.08788	0.13312	-0.09152	-0.41852	-0.10026	0.562744	-0.10972	-0.59372	-0.11398	0.794016	-0.11731	0.498992
a lot of study	0.013	0.011336	0.013572	0.015678	0.010985	0.231985	0.01144	-0.31556	0.012532	0.675532	0.013715	-0.47029	0.014248	0.922248	0.014664	0.639626
a lot of work	0.031	0.027032	0.032364	0.037386	0.026195	0.247195	0.02728	-0.29972	0.029884	0.692884	0.032705	-0.4513	0.033976	0.941976	0.034968	0.661262
aah	0.701	0.611272	0.731844	0.845406	0.592345	0.813345	0.61688	0.28988	0.675764	1.338764	0.739555	0.255555	0.768296	1.676296	0.790728	1.466602
abandoned person	0.838	0.730736	0.874872	1.010628	0.70811	0.92911	0.73744	0.41044	0.807832	1.470832	0.88409	0.40009	0.918448	1.826448	0.945264	1.631276
abandonment	-0.243	-0.211896	-0.253692	-0.29306	-0.20534	0.015665	-0.21384	-0.54084	-0.23425	0.428748	-0.25637	-0.74037	-0.26633	0.641672	-0.2741	0.331914
abase	-0.214	-0.186608	-0.223416	-0.25808	-0.18083	0.04017	-0.18832	-0.51532	-0.2063	0.456704	-0.22577	-0.70977	-0.23454	0.673456	-0.24139	0.366772
abash	-0.243	-0.211896	-0.253692	-0.29306	-0.20534	0.015665	-0.21384	-0.54084	-0.23425	0.428748	-0.25637	-0.74037	-0.26633	0.641672	-0.2741	0.331914
abashed	-0.24	-0.20928	-0.25056	-0.28944	-0.2028	0.0182	-0.2112	-0.5382	-0.23136	0.43164	-0.2532	-0.7372	-0.26304	0.64496	-0.27072	0.33552
abashment	-0.26	-0.22672	-0.27144	-0.31356	-0.2197	0.0013	-0.2288	-0.5558	-0.25064	0.41236	-0.2743	-0.7583	-0.28496	0.62304	-0.29328	0.31148
abate	0.757	0.660104	0.790308	0.912942	0.639665	0.860665	0.66616	0.33916	0.729748	1.392748	0.798635	0.314635	0.829672	1.737672	0.853896	1.533914
abatement	0.727	0.633944	0.758988	0.876762	0.614315	0.835315	0.63976	0.31276	0.700828	1.363828	0.766985	0.282985	0.796792	1.704792	0.820056	1.497854
abbreviate	-0.121	-0.105512	-0.126324	-0.14593	-0.10225	0.118755	-0.10648	-0.43348	-0.11664	0.546356	-0.12766	-0.61166	-0.13262	0.775384	-0.13649	0.478558
abdicate	0.801	0.698472	0.836244	0.966006	0.676845	0.897845	0.70488	0.37788	0.772164	1.435164	0.845055	0.361055	0.877896	1.785896	0.903528	1.586802
abdominal pregnancy	0.715	0.62348	0.74646	0.86229	0.604175	0.825175	0.6292	0.3022	0.68926	1.35226	0.754325	0.270325	0.78364	1.69164	0.80652	1.48343
abduction	-0.24	-0.20928	-0.25056	-0.28944	-0.2028	0.0182	-0.2112	-0.5382	-0.23136	0.43164	-0.2532	-0.7372	-0.26304	0.64496	-0.27072	0.33552
aberrance	0.936	0.816192	0.977184	1.128816	0.79092	1.01192	0.82368	0.49668	0.902304	1.565304	0.98748	0.50348	1.025856	1.933856	1.055808	1.749072
aberrancy	0.911	0.794392	0.951084	1.098666	0.769795	0.990795	0.80168	0.47468	0.878204	1.541204	0.961105	0.477105	0.998456	1.906456	1.027608	1.719022
aberrant	0.882	0.769104	0.920808	1.063692	0.74529	0.96629	0.77616	0.44916	0.850248	1.513248	0.93051	0.44651	0.966672	1.874672	0.994896	1.684164
abhor	0.651	0.567672	0.679644	0.785106	0.550095	0.771095	0.57288	0.24588	0.627564	1.290564	0.686805	0.202805	0.713496	1.621496	0.734328	1.406502
abhorrence	0.651	0.567672	0.679644	0.785106	0.550095	0.771095	0.57288	0.24588	0.627564	1.290564	0.686805	0.202805	0.713496	1.621496	0.734328	1.406502
abhorrent	0.712	0.620864	0.743328	0.858672	0.60164	0.82264	0.62656	0.29956	0.686368	1.349368	0.75116	0.26716	0.780352	1.688352	0.803136	1.479824
abience	0.677	0.590344	0.706788	0.816462	0.572065	0.793065	0.59576	0.26876	0.652628	1.315628	0.714235	0.230235	0.741992	1.649992	0.763656	1.437754
ability	-0.162	-0.141264	-0.169128	-0.19537	-0.13689	0.08411	-0.14256	-0.46956	-0.15617	0.506832	-0.17091	-0.65491	-0.17755	0.730448	-0.18274	0.429276
ability communicate	-0.05	-0.0436	-0.0522	-0.0603	-0.04225	0.17875	-0.044	-0.371	-0.0482	0.6148	-0.05275	-0.53675	-0.0548	0.8532	-0.0564	0.5639

〈Figure 4〉 Output results of wellness polarity values

4. Experiment and Results

4.1 Verification Method

In order to verify if the suggested method is appropriate to build WellnessWordNet, its performance was analyzed as follows. First, the gender and age of the respondents were determined, and then the levels of stress, depression, or fatigue (only one of the three) were measured on a scale of one to seven (1: strongly disagree, 7: strongly agree). Then, respondents were asked to pick one type (A or B) to answer. No hint was provided about the differences between the A and B experiments.

For those who picked experiment A, an accredited SWB measurement questionnaire was suggested. For those who picked experiment B, the following request was made: “Please pick one sentence that best describes your state of mind for today”, after which multiple options appeared formatted as “I am/feel _______ today”. The respondents then chose options pertinent to their current state. Then, the system checked if the selection could be found in WellnessWordNet. Coverage of WellnessWordNet was ensured this way.

If the word existed in WellnessWordNet, the polarity of the word for each SWB state was extracted. Three types of values were extracted: those not considering age and gender, those considering gender only, and those considering age only. For multiple answers, the polarity values were averaged. As the next phase, similarities between the average values of polarity and Subjective well-being (declared by respondents) were calculated for accuracy.

Three methods were compared as follows:

- Method 1: Estimating SWB state using an accredited SWB measurement method;
- Method 2: Estimating SWB state using Sensitivity of SenticNet; and
- Method 3: Estimating SWB state using WellnessWordNet developed for this study.

These three methods were compared to see which was most similar to the SWB state (as declared by respondents on a scale of 1 to 7), thereby revealing which one performed best. RMSE was the measure used for this purpose.

4.2 Data Collection

For purposes of the experiment, stratified sampling by gender and age was used to create our samples. An online survey was conducted for 11 days (from April 29, 2016 to May 9, 2016) by a survey agency targeting men and women in their 20s to 50s (academic background: college level or higher). Then, respondents were assigned almost equally to the two experiments (n = 502 for Type A, n = 505 for Type B, n = 1,007 in total), after their SWB state (depression, mental fatigue, and stress) had been determined. Demographic information about participants is provided in Table 9, in which the A group was used for testing Method 1 and the B group for assessing Methods 2 and 3.

〈Table 9〉 General characteristics of participants

Characteristics		A group (n = 502)	B group (n = 503)
		n (%)	n (%)
Gender	Male	252 (50.1%)	253 (50.1%)
	Female	250 (49.9%)	252 (49.9%)
Age group	20–29	125 (25.0%)	126 (25.0%)
	30–39	125 (25.0%)	126 (25.0%)
	40–49	126 (25.0%)	127 (25.0%)
	Over 50	126 (25.0%)	126 (25.0%)
Education level	College	62 (12.4%)	60 (11.9%)
	Bachelor's degree	369 (73.5%)	391 (77.4%)
	Master's and above	71 (14.1%)	54 (10.7%)
Subjective well-being	Depression	168(33.5%)	168 (33.4%)
	Fatigue	167(33.3%)	168 (33.4%)
	Stress	167(33.3%)	169 (33.6%)

4.3 Results

Table 10 indicates the results of the performance comparison. RMSE was used for measuring performance. In the cases of depression and fatigue, Method 3 with gender had the lowest RMSE (the optimized model). In the case of stress, Method 3 with age showed the lowest RMSE. When all the facts are considered, therefore, Method 3 is superior to Method 1 or Method 2. This means that if it is possible to identify age or gender, performance will be enhanced. Accordingly, WellnessWordNet, which includes gender and age, proved to be more unrestrained and accurate than ordinary questionnaires, even compared to SenticNet.

〈Table 10〉 Performance comparison by RMSE

Method	Depression	Fatigue	Stress
Method 1	23.61	22.45	16.55
Method 2	42.99	42.99	42.99
Method 3	24.37	23.70	27.36
Method 3 (gender)	14.83	13.49	17.21
Method 3 (age)	20.22	20.46	13.81

5. Discussion and Conclusion

5.1 Academic Implications

This study is academically significant in that we adopted a vocabulary/dictionary approach to measure SWB in human subjects for the first time in history. In previous research, SWB measurement has largely been reliant on ordinary questionnaires or measuring equipment that must be attached to the human body. The ordinary questionnaire method can be precise, but having to answer multiple questions one by one is tedious, restricting, and not kinetic, making it unsuitable for SWB measurement. Attaching a certain measuring device to the body is kinetic, but expensive. In some countries, SWB-related indexes are available (Diener, 2000), but the results are based on averaged values for a country as a whole; thus, they are not appropriate to use as a proxy for individuals due to inaccuracy. The method suggested in this study, however, is an unrestrained and more accurate way of measuring SWB.

Second, the combination of sentic computing and sentiment analysis from positive psychology is a welcome contribution to SWB research. Sentic computing has been utilized in multiple areas including social media marketing, Sentic chat, Sentic Avatar, and E-Health systems (Cambria and Hussain, 2012). For example, Sentic PROMs allow patients to assess their health status and health care experience in a semi-structured way (Cambria et al., 2012; Cambria et al., 2011; Cambria, 2016). The application is useful for patients who want to express their opinions and feelings in free text form, especially if they are driven by particularly positive or negative emotions. This approach, however, becomes less accurate and less plausible if the knowledge base describing human sentiments or psychological well-being state is insufficient. Thus, accuracy must be improved before the application can be useful in a commercial setting. In this study, we build a SWB analysis dictionary in order to link Sentic computing to subjective well-being research.

Third, application of a sentiment dictionary was extended to the field of wellness in this study, and demographic characteristics were reflected in the results, unlike in previous studies. The sentiment vocabularies on wellness used herein include better sentiment features for certain areas than existing sentiment dictionaries that only provide general definitions of sentiment words. Thus, our system can judge sentiments more efficiently and effectively. As WellnessWordNet also reflects individual human characteristics, it can be recommended for different services to improve well-being.

5.2 Practical Implications

More and more researchers argue that rather than GDP, SWB should be valued most as an index of life. This is because the SWB construct, which measures satisfaction and happiness, includes aspects of life beyond material conditions. Most methods to measure SWB, however, have relied on ordinary questionnaires, the multiple questions of which are highly likely to tire respondents and restrict their responses.

In this study, we suggested an unrestrained way of measuring subjective well-being. The method stipulated in this study can be used to realize a describable SWB service as well; using only a voice recognition device, psychological well-being can be measured in a natural and unrestrained manner in everyday life. Use of this service may lead to improved well-being.

5.3 Limitations

First, the performance of WellnessWordNet should be tested in a real-life setting, not a laboratory. In real life, typos, word spacing errors, slang, or jargon may appear in respondents' answers as they express their state of mind. This may complicate the measurement of SWB via WellnessWordNet.

Second, the characteristics of respondents (users) should be more diversified. Currently, only age and gender are included as demographic factors, but subjective well-being can be influenced by other factors as well, such as income level or religious faith. If such life-influential factors are added, the accuracy of the system will be further improved. In this study, only a single factor was reflected for each case; for instance, even if two factors, gender and age, are included, WellnessWordNet succeeded in deriving SWB estimates based on anger + gender or anger + age, failing to reflect the estimate from anger + gender + age. Future research may rectify this problem.

5.4 Conclusion

This study was conducted to demonstrate how WellnessWordNet can be utilized to measure SWB state. To that end, we focused on SenticNet, an existing sentiment dictionary, from which we extracted sentiment values highly related to subjective well-being. For the three most popular states of SWB, depression, fatigue, and stress, we derived statistically meaningful regression equations and estimated wellness polarity values. Including these newly-obtained estimates, the polarity values were converted into XML format, generating WellnessWordNet, an extended version of SenticNet, which was able to deduce SWB state better than other systems. In addition, the experiments demonstrated that the system performed well when individual characteristics such as gender or age are reflected. This study is meaningful for both academic and practical purposes. Future studies should complement our findings by addressing the limitations mentioned earlier and reflecting more various user characteristics to increase accuracy of individual SWB estimates. More diversified follow-up studies are therefore needed.

References

Ashimura, K., P. Baggia, F. Burkhardt, A. Oltramari, C. Peter and E. Zovato, *Vocabularies for EmotionML*, 2012. http://www.w3.org/TR/2012/NOTE-emotion-voc-20120510.

Bradley, M. M. and P. J. Lang, “Affective Norms for English Words (ANEW): Stimuli, Instruction Manual and Affective Ratings,” *Technical report*, the center for research in psychophysiology, University of Florida, 1999.

Cambria, E., T. Benson, C. Eckl and A. Hussain, “Sentic PROMs: Application of sentic computing to the development of a novel unified framework for measuring health-care quality,” *Expert Systems with Applications*, Vol.39, No.12(2012), 10533~10543.

Cambria, E. and A. Hussain, *Sentic computing: Techniques, tools, and applications,* Vol. 2, Springer Science & Business Media, 2012.

Cambria, E., A. Livingstone and A. Hussain, “The hourglass of emotions. In Cognitive Behavioural systems,” *Springer Berlin Heidelberg*, 2012, 144~157.

Cambria, E, T. Mazzocco, A. Hussain and C. Eckl, “Sentic medoids: Organizing affective common sense knowledge in a multi-dimensional vector space,” *In International Symposium on Neural Networks, Springer-Verlag, Berlin Heidelberg*

(2011), 601~610.

Cambria, E., D. Olsher and K. Kwok, "Sentic activation: A twolevel affective common sense reasoning framework In: Proce," *edings of AAAI*(2012), 186~192.

Cambria, E., "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, Vol.31, No.2(2016), 102~107.

Charlesworth, E. A., B. J. Williams and P. E. Baer, "Stress management at the worksite for hypertension: compliance, cost-benefit, health care and hypertension-related variables," *Psychosomatic medicine*, Vol.46, No.5(1984), 387~397.

Diener, E., "Subjective well-being: The science of happiness and a proposal for a national index," *American Psychologist*, Vol.55, No.1(2000), 34.

Diener, E. D., R. A. Emmons, R. J. Larsen and S. Griffin, "The satisfaction with life scale," *Journal of Personality Assessment*, Vol.49, No.1(1985), 71~75.

Dolan, P., T. Peasgood and M. White, "Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being," *Journal of Economic Psychology* Vol.29, No.1(2008), 94~122.

Dupuy, H. J., *Self-representations of general psychological well-being of American adults,* In American Public Health Association Meeting, Los Angeles (1978).

Esuli, A. and F. Sebastiani, Sentiwordnet: A publicly available lexical resource for opinion mining. *In Proceedings of LREC*, Vol.6(2006), 417~422.

Havighurst, R. J., B. L. Neugarten and S. S. Tobin, "The measurement of life satisfaction," *Journal of Gerontology,* Vol.16, No.2(1961), 134~143.

Hung, C. and H. K. Lin, "Using objective words in SentiWordNet to improve word-of-mouth sentiment classification," *IEEE Intelligent Systems*, Vol.28, No.2(2013), 47~54.

Kozma, A. and M. J. Stones, "Social desirability in measures of subjective well-being: A systematic evaluation," *Journal of Gerontology*, Vol.42, No.1(1987), 56~59.

Krupp, L. B., N. G. LaRocca, J. Muir-Nash and A. D. Steinberg, "The fatigue severity scale: application to patients with multiple sclerosis and systemic lupus erythematosus," *Archives of Neurology*, Vol.46, No.10(1989), 1121~1123.

Lawton, M. P. "The Philadelphia Geriatric Center Morale Scale: A revision," *Journal of Gerontology*, Vol.30, No.1(1975), 85~89.

Lee, M. S. and M. K. Rhee, "A development of Korea depression scale," *Journal of Korean Neuropsychiatric Association*, Vol.42, No.4 (2003), 492~506.

Liu, Q. and S. Li, "Word similarity computing based on How-net," *Computational Linguistics and Chinese Language Processing*, Vol.7, No.2(2002), 59~76.

McBride, M., "Relative-income effects on subjective well-being in the cross-section," *Journal of Economic Behavior & Organization*, Vol.45, No.3(2001), 251~278.

Pezzuti, L., E. Longobardi, S. Rossetti, E. Bartolini and G. Natale, "The Relation between the Theory of Mind and Socio-Emotional Functioning in a Sample of Older Adults," *Journal of Neurology and Psychology*, Vol.3, No.2(2015), 1~7.

Sánchez-Rada, J. F. and C. A. Iglesias, "Onyx: A linked data approach to emotion representation," *Information Processing & Management,* Vol.52, No.1(2016), 99~114.

Shmotkin, D., "Subjective well-being as a function of age and gender: A multivariate look for differentiated trends. Social Indicators Research," Vol. 23, No.3(1990), 201~230.

Smith, J., M. Borchelt, H. Maier and D. Jopp, "Health and well–being in the young old and oldest old," *Journal of Social Issues*, Vol.58, No.4(2012), 715~732.

Spielberger, C. D., "Manual for the state-trait anger expression scale (STAXI)," *Odessa, FL: Psychological Assessment Resources*, 1988.

Stone, P. J. and E. B. Hunt, "A computer approach to content analysis: studies using the general inquirer system," *In Proceedings of the May 21-23, Spring Joint Computer Conference*, (1963), 241-256.

Strapparava, C. and A. Valitutti, "WordNet Affect: an Affective Extension of WordNet," *In LREC*, Vol. 4(2004), *1083~1086*.

Watson, D., L. A. Clark and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the PANAS scales," *Journal of Personality and Social Psychology*, Vol.54, No.6(1988), 1063~1070.

Xia, Y., X. Li, E. Cambria and A. Hussain, "A Localization Toolkit for Sentic Net," *In 2014 IEEE International Conference on Data Mining Workshop (ICDMW),* 403~408.

Zhang, X., X. Chen, G. Ran and Y. Ma, "Adult children's support and self-esteem as mediators in the relationship between attachment and subjective well-being in older adults," *Personality and Individual Differences*, Vol.97(2016), 229~233.

Appendix A. Examples of created XML codes in the WellnessWordNet

(1) **hateful** [Sensitivity: 0.596]

Sentic API sample (Source: http://sentic.net/api)


```xml
<rdf:Description rdf:about="http://sentic.net/api/en/concept/hateful">
    <rdf:type rdf:resource="http://sentic.net/api/concept"/>
    <text xmlns="http://sentic.net/api">hateful</text>
    <sensitivity xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.596</sensitivity>
    <wellnesswordnet xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.596</wellnesswordnet>
    <depression xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.520</depression>
    <depression_male xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.504 </depression_male>
    <depression_female xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.725 </depression_female >
    <depression_lt50 xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.524</depression_lt50 >
    <depression_gt50 xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.197</depression_gt50 >
    <fatigue xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.622</fatigue>
    <fatigue_male xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.575</fatigue_male >
    <fatigue_female xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">1.238 </fatigue_female >
    <fatigue_lt50 xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.629</fatigue_lt50 >
    <fatigue_gt50 xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.145 </fatigue_gt50 >
    <stress xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.719 </stress>
    <stress_male xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.653 </stress_male >
    <stress_female xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">1.561 </stress_female >
    <stress_lt30 xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.672 </stress_lt30 >
    <stress_gt30 xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">1.296 </stress_gt30 >
</rdf:Description>
```


(2) **gloomy** [Sensitivity: -0.471]

Sentic API sample (Source: http://sentic.net/api)


```xml
<rdf:Description rdf:about="http://sentic.net/api/en/concept/gloomily">
    <rdf:type rdf:resource="http://sentic.net/api/concept"/>
    <text xmlns="http://sentic.net/api">gloomily</text>
    <sensitivity xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">-0.471</sensitivity>
    <wellnesswordnet xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.471</wellnesswordnet>
    <depression xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.411 </depression>
    <depression_male xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.398 </depression_male>
    <depression_female xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.619 </depression_female >
    <depression_lt50 xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.414 </depression_lt50 >
    <depression_gt50 xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.087 </depression_gt50 >
    <fatigue xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.492 </fatigue>
    <fatigue_male xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.454 </fatigue_male >
    <fatigue_female xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">1.117 </fatigue_female >
    <fatigue_lt50 xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.497 </fatigue_lt50 >
    <fatigue_gt50 xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.013 </fatigue_gt50 >
    <stress xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.568 </stress>
    <stress_male xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.516 </stress_male >
    <stress_female xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">1.424 </stress_female >
    <stress_lt30 xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.531 </stress_lt30 >
    <stress_gt30 xmlns="http://sentic.net/api" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">1.155 </stress_gt30 >
</rdf:Description>
```

국문요약

웰니스워드넷: 비정형데이터와 상황적 공부정성에 기반하여 주관적 웰빙 상태를 무구속적으로 모니터링하기 위한 워드넷 개발*

송영은**·남수현**·권오병***

주관적 웰빙 서비스(subjective well-being service)는 Wellness IT의 주요 서비스이며 개인의 주관적 웰빙 상태를 무구속적이고 비용 효율적으로 측정하는 방법이 중요하다. 이를 위해 감성어휘사전을 활용할 수 있으나 감성어만으로 주관적 웰빙 상태를 측정할 수는 없으며 웰니스 어휘 사전이 별도로 구축될 필요가 있다. 더욱이 기존의 감성어휘사전은 동일한 감정어에 대해 한가지만의 감성값을 제공함으로써 그 용어를 사용한 사람의 특징에 따라 감성값이 변경될 수 있다는 점을 간과하고 있다. 따라서 본 연구의 목적은 현존하는 감성어휘사전 중에서 표현력이 가장 뛰어난 SenticNet을 기반으로 하여 SenticNet에서 제공하는 정보를 통해 스트레스, 우울, 분노, 행복감 등 웰니스 상태를 추정한 결과를 추가한 WellnessWordNet 을 개발하는 것이다. 또한 실제 사람들을 대상으로 WellnessWordNet 에 근거한 웰니스 상태 추정 정확도를 검증해 보았다. 본 논문의 독창성은 WellnessWordNet 웰니스 상태 언어에 대한 값을 제공할 뿐더러, 성별이나 연령과 같은 사람의 특성에 따라 다른 감성값을 제공하는 최초의 감성어휘사전이라는 것이다.

주제어 : 주관적 웰빙; 감성분석; 상황적 공부정성; 비정형데이터; 센틱넷

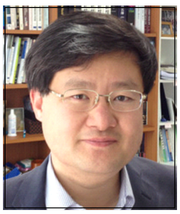
논문접수일 : 2016년 7월 20일 논문수정일 : 2016년 8월 10일 게재확정일 : 2016년 8월 11일
원고유형 : 일반논문 교신저자 : 권오병

* 이 논문은 2016년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.R0126-15-1007, 맞춤형 개인 행복 증진을 위한 큐레이션 커머스용 글로벌 오픈 마켓 구축 기술 개발)
** 경희대학교 일반대학원 경영학과
*** 교신저자 : 권오병
경희대학교 경영대학
26 Kyungheedae-ro, Dongdaemun-gu, Seoul 130-701, Korea
Tel: +82-2-961-2148, Fax: +82-2-961-0515, E-mail: obkwon@khu.ac.kr

저 자 소 개

송 영 은
현재 경희대학교 일반대학원 경영학과에서 빅데이터 경영전공 석사과정에 재학 중이며, 목원대학교 정보컨설팅학과에서 학사학위를 취득한 바 있다. 그리고 Caitech연구소에서 개인 행복 증진을 위한 큐레이션 커머스 프로젝트에 참여 중이다. 주요 관심 분야는 텍스트 마이닝에 기반한 빅데이터 분석, 감성 분석이다.

남 수 현
경희대학교 경영학과에서 학사학위를 취득하였으며, 현재 경희대학교 일반대학원 경영학과에서 빅데이터 경영 전공으로 석사과정 재학 중이다. 경희대학교 BK21 플러스 데이터 과학에 기반한 경영 전문 연구인력 양성팀 연구원 재직 및 개인 행복 증진을 위한 큐레이션 커머스 프로젝트에 참여 중이다. 주요관심분야는 Data Mining, Big data analysis, Text Mining, Algorithm Implementation 등이다.



권 오 병
현재 경희대학교 경영학과 교수로 재직 중이다. 1988년 서울대학교 경영학과(경영학사), 1990년 한국과학기술원 경영과학과(공학석사), 1995년 한국과학기술원 경영과학과(공학박사)를 졸업하였다. 2001년~2002년에는 카네기멜론대학 전산학부에서 방문과학자로 근무한 바 있으며 2009년~2011년에는 샌디에고주립대학 경영정보학과의 겸직교수로 재직한 바 있다. 관심분야는 빅데이터분석, 사물인터넷, 의사결정지원시스템 등이다.