

Development and Evaluation of an Obesity Ontology for Social Big Data Analysis

Ae Ran Kim, MSN, RN^{1,2}, Hyeoun-Ae Park, PhD, RN, FAAN^{1,3}, Tae-Min Song, PhD⁴

¹College of Nursing & Systems Biomedical Informatics Research Center, Seoul National University, Seoul, Korea; ²Department of Nursing, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; ³Research Institute of Nursing Science, Seoul National University, Seoul, Korea; ⁴Korea Institute for Health and Social Affairs, Sejong, Korea

Objectives: The aim of this study was to develop and evaluate an obesity ontology as a framework for collecting and analyzing unstructured obesity-related social media posts. **Methods:** The obesity ontology was developed according to the ‘Ontology Development 101’. The coverage rate of the developed ontology was examined by mapping concepts and terms of the ontology with concepts and terms extracted from obesity-related Twitter postings. The structure and representative ability of the ontology was evaluated by nurse experts. We applied the ontology to the density analysis of keywords related to obesity types and management strategies and to the sentiment analysis of obesity and diet using social big data. **Results:** The developed obesity ontology was represented by 8 superclasses and 124 subordinate classes. The superclasses comprised ‘risk factors,’ ‘types,’ ‘symptoms,’ ‘complications,’ ‘assessment,’ ‘diagnosis,’ ‘management strategies,’ and ‘settings.’ The coverage rate of the ontology was 100% for the concepts and 87.8% for the terms. The evaluation scores for representative ability were higher than 4.0 out of 5.0 for all of the evaluation items. The density analysis of keywords revealed that the top-two posted types of obesity were abdomen and thigh, and the top-three posted management strategies were diet, exercise, and dietary supplements or drug therapy. Positive expressions of obesity-related postings has increased annually in the sentiment analysis. **Conclusions:** It was found that the developed obesity ontology was useful to identify the most frequently used terms on obesity and opinions and emotions toward obesity posted by the general population on social media.

Keywords: Obesity, Social Media, Knowledge, Data Analysis, Terminology

Submitted: December 22, 2016

Revised: June 18 2017

Accepted: June 28, 2017

Corresponding Author

Hyeoun-Ae Park, PhD, RN, FAAN

College of Nursing, Seoul National University, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea. Tel: +82-2-740-8827, E-mail: hapark@snu.ac.kr

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2017 The Korean Society of Medical Informatics

1. Introduction

The introduction of information and communication technologies in healthcare has made it possible to collect medical and health information on patients through Electronic Health Records, digital imaging, and smart sensors [1,2]. This has led to an explosion of data on healthcare [2,3], and such data have also accumulated in cyberspace due to the development of smartphones and Internet services [3]. People search and share health-related information [4] and establish connections with people who have similar health conditions on social media [5]. Comments posted on social media can provide real-world observations of public awareness of obesity [6].

The increasing use of social media has prompted researchers to collect social media messages and analyze their meaning using social media analytics [3]. Social media analytics involves extracting and analyzing the information in messages posted on social media, such as Twitter or Facebook [2]. For instance, a study was conducted to predict suicide using social media messages presumed to be written by teenagers and to provide tailored programs for suicide prevention [7].

Text mining is one of the methods used to analyze unstructured text data from social media [2,8,9]. The accurate analysis of text data requires an understanding of the precise meaning of the terms included in documents [9] and the development of an analytic framework for data collection and analysis. An ontology is a computer-interpretable knowledge model for formalizing and representing shared concepts of the subject of interest [10]. An ontology can provide a framework to extract information from data and derive knowledge from information [1]. An ontology defines concepts related to topics, relationships of concepts, and the vocabulary to represent certain domains [11].

Health information related to obesity is often exchanged through social media. Obesity is a risk factor for several chronic diseases, such as hypertension, type II diabetes, hyperlipidemia, and coronary artery disease [12]. In Korea, the prevalence of obesity among adults aged 19 years and older was 32.5% in 2013 (37.6% in males and 27.5% in females) [13], and the medical costs of diagnosing and treating obesity are increasing rapidly [14] with a considerable associated socioeconomic burden.

The collection and analysis of social big data related to

obesity could improve our understanding of the public awareness of obesity, coping strategies, and the related social phenomena. Therefore, we developed and evaluated an obesity ontology as a framework for collecting and analyzing big data on social media sites in this study.

II. Methods

This study was carried out in the three stages presented in Figure 1.

1. Development

We developed an obesity ontology according to the ‘Ontology Development 101’ described by Noy and McGuinness [11]. We used this because it provides very concrete and applicable guidance on the required tasks in each step. Its iterative design allows ontology developers revise an ontology easily [15,16].

1) Determining the domain and scope of the ontology

We chose the domain and scope of our ontology based on social big data in order to answer the following questions: “How do individuals perceive the factors contributing to obesity?” “What are the sociodemographic characteristics of obese persons?” “How do individuals and healthcare providers evaluate the types and degree of obesity?” “What types of obesity exist?” “What are the results of obesity?” and “How do individuals and healthcare providers treat and prevent obesity?”

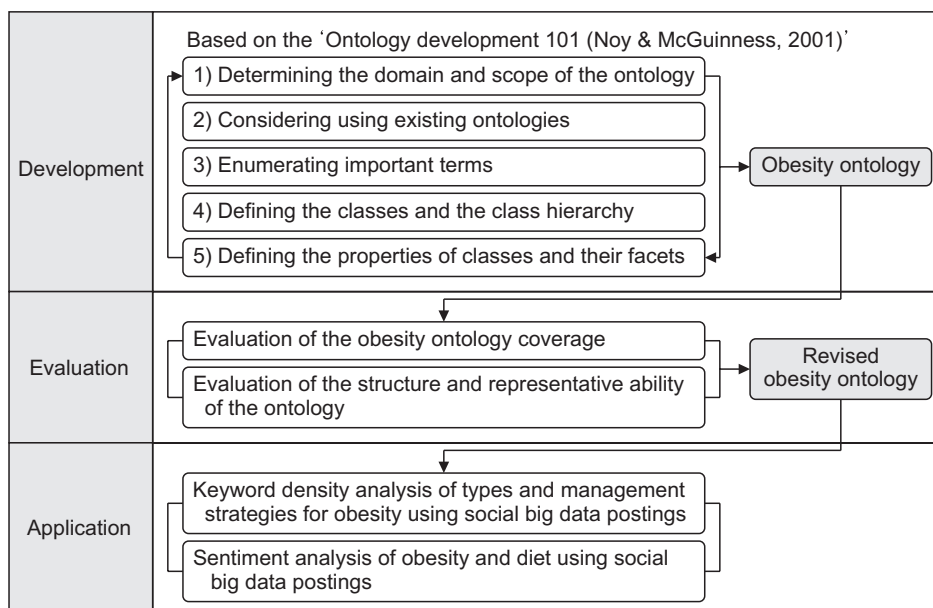


Figure 1. The research process.

2) Considering the reuse of existing ontologies

We searched publications using the PubMed, EBSCO, and Google Scholar databases with the keyword ‘obesity ontology.’ We then reviewed the purpose of development and the extracted concepts of the retrieved ontologies to consider their utilization.

We identified five obesity-related ontologies [17-21] whose purposes and domains are presented in Table 1. Two of the ontologies [17,18] were developed for the diagnosis and evaluation of obesity, but they did not address its treatment or prevention. The obesity ontologies for intelligent e-therapy [19] and for application in the mobile-device domain [20] included not only evaluations but also interventions for obesity, but they excluded medical treatment. Finally, an ontology for supporting a knowledge-based infrastructure of childhood obesity [21] included generic and customized preventive recommendations and several risk factors of childhood obesity, but did not cover adult obesity.

Since each of the retrieved ontologies dealt with only part of obesity-related concepts, it was necessary to develop a new ontology to integrate the following concepts: factors contributing to obesity, sociodemographic characteristics of obese persons, evaluation methods, types, signs, symptoms, complications, medical treatment or interventions by health-care providers, and self-management strategies of obesity.

3) Enumerating important ontology terms

We enumerated the important terms related to the domains and scope of our ontology. The core concepts and terms related to obesity were extracted from the clinical practice guidelines on obesity published by the National Institute for Health and Clinical Excellence (NICE) [22] and the Korean Society for the Study of Obesity (KSSO) [23]. We searched obesity-related social media postings on Internet blogs,

Twitter, and Facebook from January 2011 to June 2014 by combining ‘obesity’ and obesity-related keywords representing the following domains: obesity risk factors, sociodemographic characteristics, evaluation methods, types, signs, symptoms, complications, and management strategies. We also extracted additional terms from the retrieved postings that are not included in the guidelines.

4) Defining the classes and the class hierarchy

The extracted terms were enumerated using a top-down method [11], starting from the most-general concept to more detailed concepts, to define the common concepts as classes and order them hierarchically.

5) Defining the properties of classes and their facets

The properties of classes and their facets, such as value types, allowed values, cardinality, and other features of the values, were defined to describe the internal structure of the extracted concepts.

The obesity ontology was developed by defining the relationships of the classes. The terminology was developed by linking terms to the concepts of the obesity ontology and by identifying synonyms for each concept.

2. Evaluation

1) Evaluation of the obesity ontology coverage

The coverage of the ontology developed in this study was evaluated by mapping the concepts and terms of the ontology with those extracted from the obesity-related Twitter postings. We performed a Twitter search using the combinations of keywords related to obesity (e.g., ‘diet,’ ‘exercise,’ ‘sleep,’ ‘smoking,’ and ‘drinking’) from January to March 2015 and selected the postings related to obesity. After converting them into a text file, keywords related to obesity were

Table 1. Review of existing obesity-related ontologies

Study	Purpose of the ontology	Domains of the ontology
Scala et al. [17], 2012	To evaluate obesity and related comorbidities	Person, physiological profile, body structure, diet, and physical activity
Jeon and Ko [18], 2007	To implement a Bayesian network for diagnosing obesity	Obesity, obesity causes, disease, symptoms, and biodata
Zaragoza et al. [19], 2009	To be integrated into an intelligent e-therapy for obesity	Agent, evaluation, alarm, and treatment
Kim et al. [20], 2013	To apply in the mobile-device domain for obesity management	Assessment data, inferred data to represent nursing diagnoses and evaluation, and implementation
Shaban-Nejad et al. [21], 2011	To support a knowledge-based infrastructure for childhood obesity prevention	Nutrition, obesity and chronic disease, behaviors, media, and marketing

extracted by natural language processing using the KoNLP package in the R software. The frequencies of the extracted keywords were tallied, and any extracted keywords that were not included in the ontology were added to the revised obesity ontology.

2) Evaluation of the structure and representative ability of the ontology

The structure and representative ability of the revised obesity ontology were evaluated by nurse experts using seven items and ten items ontology evaluation criteria on a 5-point Likert-type scale, respectively, developed by Kehagias et al. [24]. We also analyzed the content validity index (CVI) by dividing the number of experts responding to 4 or 5 points by the total number of experts for each evaluation criteria. Three experts who had a master's or higher degree in nursing informatics and one expert who had a doctoral degree in adult health nursing with experience in education and research related to obesity participated in the evaluation. The experts were asked to assess the structure of the entire ontology and the representative ability of the superclasses of the ontology.

3. Application of the Developed Ontology in the Analysis of Obesity-Related Social Big Data

We analyzed obesity-related discourses posted on social media by utilizing the ontology developed in this study. We

collected social big data on obesity using the keywords of 'obesity' and 'diet' from 217 online news sites, 4 blogs, 2 social network services, and 11 online bulletin boards posted between January 2011 and December 2013. An automated software crawler was used to collect social big data for this study. We analyzed the keyword density to answer "How do individuals describe the types of obesity?" and "What kinds of obesity management strategies were described in social media postings?" We categorized the postings on obesity and diet into positive, neutral, and negative sentiments to answer "How individuals think and feel about the concepts of obesity and diet?"

III. Results

1. Development

1) Enumerating important terms in the ontology

We extracted 1,712 terms from the clinical practice guidelines on obesity published by NICE [22] and KSSO [23] and 396 additional terms from social media postings on obesity.

2) Defining the classes and class hierarchy

We defined the class hierarchy by creating classes for the general concepts of 'risk factors,' 'sociodemographic characteristics,' 'assessment,' 'diagnosis,' 'types,' 'symptoms and complications,' 'treatment,' and 'prevention.' We then divided

Table 2. Definition of superclasses and their hierarchy

Superclass	Definition	Depths	Number of subordinate classes	Examples of keywords
Risk factors	Factors that contribute to obesity in individuals	4	42	Risk factors, causes, contributing factors
Types	Classification of obesity in individuals based on diagnosis and assessment	2	3	Types, classification
Symptoms and complications	Symptoms or complications of obesity	3	24	Signs, symptoms, complications
Assessment	Data collection method to assess the degree and type of obesity by healthcare providers or clients	2	2	Assessment, data collection, evaluation
Diagnosis	Body measurements and test methods used to diagnose the degree and type of obesity in an individual by healthcare providers or clients	2	2	Diagnosis, identification
Management strategies	Treatment or prevention methods applied to manage obesity by healthcare providers or clients	5	73	Management, control, interventions
Settings	Environment of assessment, diagnosis, or management strategies for obesity	2	2	Settings, environment, places

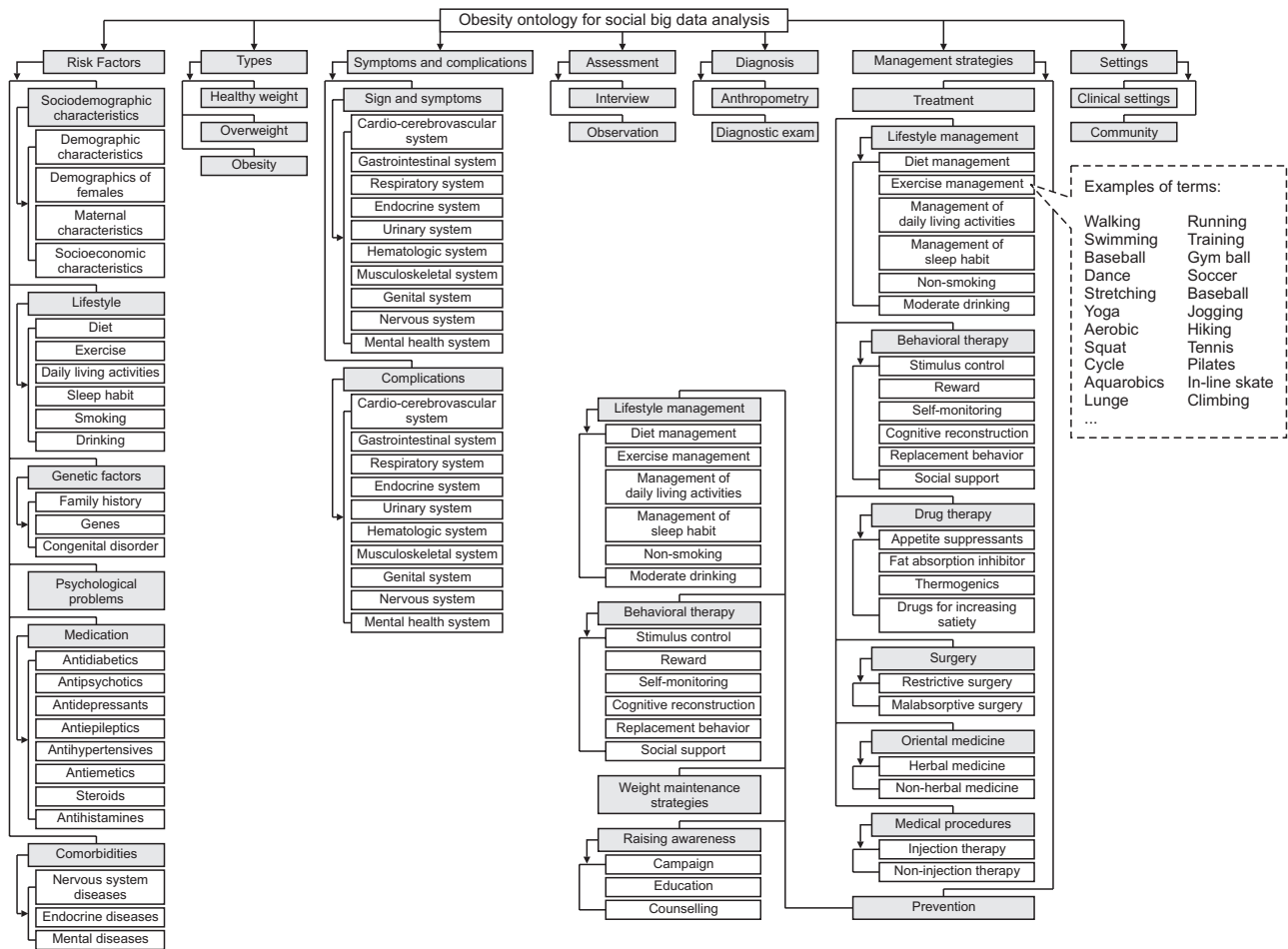


Figure 2. Superclasses and main subclasses in the obesity ontology.

Table 3. Example of properties and facets of the 'risk factor-lifestyle-diet-usual diet' class

Class	Property	Value type	Values	Cardinality
Usual diet	Preferred dietary type	SC	[high-calorie diet high-fat diet high-carbohydrate diet high-salt diet fast foods instant foods]	Multiple
	Preferred meal	ST	e.g., burger, coke, fried potatoes	Multiple
	Preferred cooking methods	SC	[pan-frying frying broiling steaming boiling blanching reducing seasoning]	Multiple
	Intake amount	SC	[low adequate excessive]	Single
	Diet regularity	SC	[regular irregular]	Single
	Average speed of consuming a meal	SC	[slow normal fast]	Single
	Average number of meals per day	SC	[<3 3 >3]	Single
	Average number of snacks per day	SC	[<3 3 >3]	Single
	Breakfast	SC	[never sometimes often always]	Single
	Dining out	SC	[never sometimes often always]	Single
	Late-night snack	SC	[never sometimes often always]	Single

these classes into specialized subclasses. We combined ‘treatment’ and ‘prevention’ concepts into the ‘management strategies’ superclass. To address the concept of the environment where assessments, diagnosis, and management strategies for obesity are conducted, the concept of ‘settings’ was added as a superclass, and ‘sociodemographic characteristics’ was moved down to a subordinate concept for ‘risk factors.’ Table 2 lists the definitions of the superclasses, depths, and number of subordinate classes. The final obesity ontology used for the collecting and analyzing obesity-related social big data comprised 7 superclasses and 148 subclasses. We present the superclasses and subclasses down to the 4th level under each superclass in Figure 2.

3) Defining the properties of the classes and their facets

We defined the properties and their facets describing the internal structure of each concept, an example of which is presented in Table 3. On average, 4.0 properties and 2.0 relations were defined within each class.

2. Evaluation

1) Coverage evaluation of the obesity ontology

In total, 520 postings were retrieved from Twitter between January and March 2015. After excluding 79 duplicated postings, we selected 441 obesity-related postings for coverage evaluation. We manually reviewed 2,981 words extracted by natural language processing and confirmed the presence of 452 obesity-relevant terms.

Examples of terms that appeared at higher frequencies were ‘obesity’ (n = 191, 43.3%), ‘exercise’ (n = 134, 30.4%), ‘diet’ (n = 50, 11.3%), ‘lower-body obesity’ (n = 41, 9.1%), ‘fat’ (n = 37, 8.4%), and ‘treatment’ (n = 35, 7.9%). The obesity ontol-

Table 4. Evaluation scores for the structure of the revised obesity ontology

Criterion	Mean (range) / CVI
Size	5.0 (5.0–5.0) / 1.0
Depth of hierarchy	5.0 (5.0–5.0) / 1.0
Breadth of hierarchy	5.0 (5.0–5.0) / 1.0
Density	4.3 (4.0–5.0) / 1.0
Balance	5.0 (5.0–5.0) / 1.0
Overall complexity	4.8 (4.0–5.0) / 1.0
Connectivity between concepts and relationships	4.0 (3.0–5.0) / 0.8

CVI: content validity index.

Table 5. Evaluation scores for the representative ability of the revised obesity ontology

Criterion	Risk factors Mean (range)/CVI	Types Mean (range)/CVI	Symptoms and complications Mean (range)/CVI	Assessment Mean (range)/CVI	Diagnosis Mean (range)/CVI	Management strategies		Settings		Overall	
						Mean (range)/CVI	Mean (range)/CVI	Mean (range)/CVI	Mean (range)/CVI	Mean (range)/CVI	Mean (range)/CVI
Match	5.0 (5.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0	5.0 (5.0–5.0) / 1.0	4.5 (3.0–5.0) / 0.8	4.5 (3.0–5.0) / 0.8	4.3 (2.0–5.0) / 0.8	5.0 (5.0–5.0) / 1.0	5.0 (5.0–5.0) / 1.0	4.7 (3.0–5.0) / 0.8		
Consistency	4.3 (3.0–5.0) / 0.8	4.8 (4.0–5.0) / 1.0	4.5 (4.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0	4.3 (3.0–5.0) / 0.8	4.8 (4.0–5.0) / 1.0	5.0 (5.0–5.0) / 1.0	5.0 (5.0–5.0) / 1.0	4.6 (4.0–5.0) / 1.0		
Clarity	4.8 (4.0–5.0) / 1.0	5.0 (5.0–5.0) / 1.0	5.0 (5.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0	5.0 (5.0–5.0) / 1.0	4.5 (3.0–5.0) / 0.8	5.0 (5.0–5.0) / 1.0	5.0 (5.0–5.0) / 1.0	4.9 (4.0–5.0) / 1.0		
Explicitness	4.8 (4.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0	5.0 (5.0–5.0) / 1.0	4.5 (3.0–5.0) / 0.8	5.0 (5.0–5.0) / 1.0	5.0 (5.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0		
Interpretability	4.8 (4.0–5.0) / 1.0	4.5 (3.0–5.0) / 0.8	4.8 (4.0–5.0) / 1.0	4.5 (3.0–5.0) / 0.8	4.8 (4.0–5.0) / 1.0	4.3 (2.0–5.0) / 0.8	4.5 (3.0–5.0) / 0.8	4.5 (3.0–5.0) / 0.8	4.6 (3.0–5.0) / 0.8		
Accuracy	4.5 (3.0–5.0) / 0.8	4.3 (3.0–5.0) / 0.8	4.0 (3.0–5.0) / 0.5	4.3 (2.0–5.0) / 0.8	4.3 (3.0–5.0) / 0.8	4.3 (3.0–5.0) / 0.8	4.5 (4.0–5.0) / 1.0	4.3 (3.0–5.0) / 0.5			
Comprehensiveness	5.0 (5.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0	4.5 (4.0–5.0) / 1.0	4.5 (4.0–5.0) / 1.0	4.5 (3.0–5.0) / 0.8	4.5 (3.0–5.0) / 1.0	4.7 (4.0–5.0) / 1.0		
Granularity	4.8 (4.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0	4.5 (4.0–5.0) / 1.0	4.5 (3.0–5.0) / 0.8	4.5 (4.0–5.0) / 1.0	4.3 (3.0–5.0) / 0.8	4.5 (4.0–5.0) / 1.0	4.5 (3.0–5.0) / 0.8			
Relevance	4.5 (4.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0	4.5 (3.0–5.0) / 0.8	4.3 (3.0–5.0) / 0.8	4.3 (2.0–5.0) / 0.8	5.0 (5.0–5.0) / 1.0	4.6 (3.0–5.0) / 0.8			
Description	4.3 (3.0–5.0) / 0.8	4.8 (4.0–5.0) / 1.0	4.8 (4.0–5.0) / 1.0	4.5 (3.0–5.0) / 0.8	4.5 (3.0–5.0) / 0.8	4.5 (3.0–5.0) / 0.8	4.8 (4.0–5.0) / 1.0	4.6 (3.0–5.0) / 0.8			

Values are presented as mean (range) / CVI. CVI: content validity index.

ogy developed in this study included all of these terms.

The 452 extracted terms were mapped to the classes (n = 85), properties (n = 64), and values (n = 380) with a 100% coverage rate for the concepts. However, our ontology included 397 of these 452 obesity-related terms, giving it a coverage rate of 87.8%. The 55 terms not mapped were synonyms for concepts of the classes (n = 3), properties (n = 5), and values (n = 47) of our ontology. These were mostly terms representing values of the ‘management strategies’ superclass in our ontology, such as various types of exercise (e.g., ‘dead lifts,’ ‘burpees,’ and ‘push-ups’). We revised the obesity ontology by adding these terms.

2) Evaluation of the structure and representative ability of the revised obesity ontology

Tables 4 and 5 list the evaluation scores for the structure and representability of the ontology, respectively. The experts rated all seven criteria of the ontology structure at 4.0 or higher (Table 4). The criterion with the lowest score was the relation between concepts, and an expert pointed out that the relationships between the ‘settings’ class and other classes were not clearly represented. We therefore revised the ontology by redefining the relationships between the ‘settings’ superclass with the ‘assessment,’ ‘diagnosis,’ and ‘management strategies’ superclasses.

All ten criteria for representability scored higher than 4.0 (Table 5). The criterion with the lowest score was ‘accuracy’ at 4.3, which indicates how well the ontology included the terms representing obesity-related concepts on social media. The scores for these superclasses were probably lower because these concepts included many medical terms representing drugs, diseases, or medical treatment interventions. In particular, the experts rated the ‘symptom and complications’ superclass as 4.0. One expert suggested dividing the ‘symptom and complications’ superclass into two separate superclasses because these are very different concepts. Based

on this suggestion we revised the ontology with 8 superclasses and 124 subclasses (Figure 3). The terminology of the revised ontology comprised 278 terms for classes (including 162 synonyms), 147 terms for properties (including 3 synonyms), and 1,737 terms for values (including 865 synonyms).

3. Application of the Developed Ontology in the Analysis of Obesity-Related Social Big Data

We collected a total of 1,207,531 obesity-related postings that posted on social media from January 2011 to December 2013.

1) Keyword density analysis

The ranks of the frequently used keywords on types of obesity and management strategies for obesity are presented in Table 6. High-frequency terms related to types of obesity were ‘abdominal fat’ (n = 16,830), ‘thigh’ (n = 15,931), and ‘lower-body obesity’ (n = 13,844). ‘Fat’ (n = 36,821) was the most frequently mentioned term related to diet. Terms related diet methods, such as ‘one-food diet,’ ‘Danish diet,’ or ‘Atkins diet’ were also mentioned frequently as diet terms. ‘Stretching’ (n = 8,469), ‘aerobic exercise’ (n = 8,387), ‘walking’ (n = 7,544), and ‘yoga’ (n = 4,779) were the high-frequency exercise terms involving activities that could be performed without any restriction of equipment or place. The terms ‘shakes’ (n = 7,807), ‘Herbalife’ (n = 7,165), and ‘Chinese medicine’ (n = 4,009) were the top-three keywords related to drug or diet supplement therapy.

2) Sentiment analysis

The results of the sentiment analysis for obesity and diet are shown in Figure 4. Neutral comments were the most common (66%), followed by positive comments (22%) and negative comments (12%) in the 1,441,939 obesity and diet related postings. The examples of the keywords from positive

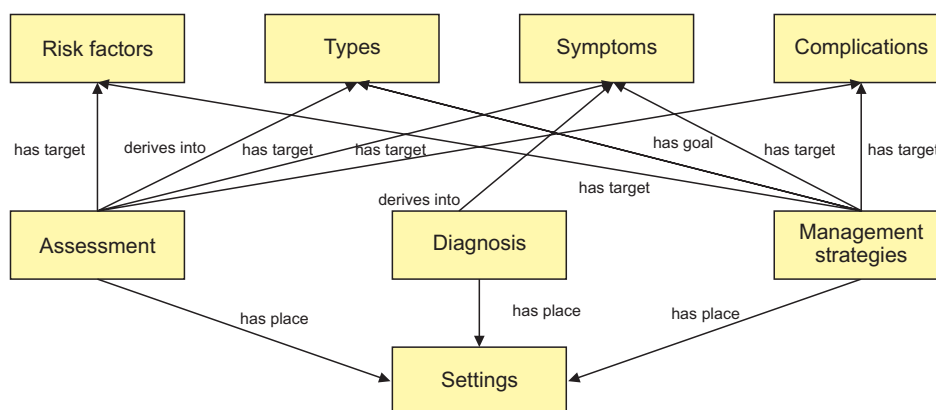


Figure 3. Revised obesity ontology for collecting and analyzing big data related to obesity.

Table 6. Results of keyword density analysis of obesity-related social big data

Rank	Types of obesity (body part)		Management strategies					
			Diet		Drug therapy or dietary supplement		Exercise	
	Keyword	Frequency	Keyword	Frequency	Keyword	Frequency	Keyword	Frequency
1	Abdominal fat	16,830	Fat	36,821	Shakes	7,807	Stretching	8,469
2	Thigh	15,931	Eggplant	21,470	Herbalife	7,165	Aerobic exercise	8,387
3	Lower-body obesity	13,844	Protein	20,569	Chinese medicine	4,009	Walking	7,544
4	Face	12,577	Vegetables	18,659	Aloe gel	2,303	Yoga	4,779
5	Abdominal obesity	12,394	Fruit	15,219	Nutraceutical	922	Jump rope	4,582
6	Waist	10,695	Diet therapy	14,339	Anti-obesity drug	547	Muscular exercise	3,517
7	Hip	5,975	Vitamins	11,814	Appetite suppressant	381	Cycling	2,863
8	Calf	4,437	Salad	9,415	Sibutramine	198	Swimming	2,270
9	Forearm fat	2,967	Calorie	7,096	Phentermine	10	Jogging	1,392
10	Upper-body obesity	1,161	Milk	3,139	Orlistat	6	Climbing stairs	742

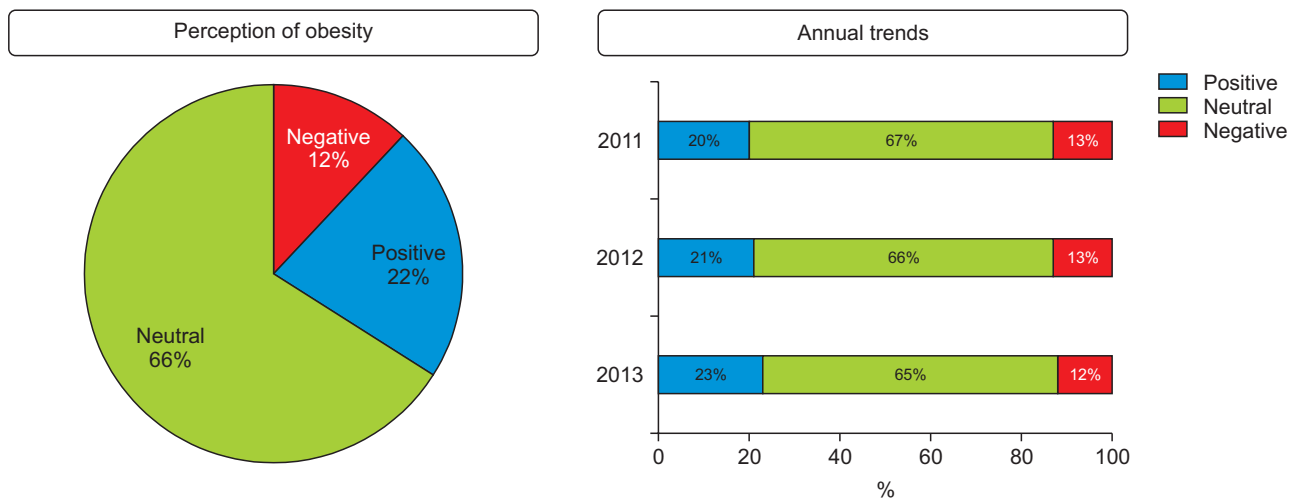


Figure 4. Results of sentiment analysis of obesity-related social big data (n = 1,441,939).

comments were 'to prevent obesity' and 'to manage obesity,' 'successful diet' and 'healthy diet,' and 'helpful exercise' and 'effective exercise.' The examples of the keywords from negative comments were 'skinny fat' and 'super-obesity,' 'failed diet' and 'excessive diet,' and 'exhausted from exercise' and 'no time for exercise.' It was found that the positive perception tend to increase annually during the 3-year analysis period.

IV. Discussion

In this study we developed and evaluated an obesity ontology as a framework for collecting and analyzing obesity-related social big data. The obesity ontology developed in this study comprises 8 superclasses and 124 subclasses with two to

five levels deep. The terms representing the ontology classes comprises 1,121 preferred terms with 1,030 synonyms.

The superclasses of our ontology integrate all of the concepts in the existing ontologies on obesity [17-21] such as risk factors [17,18,21], symptoms or complications [18], diagnosis [19], assessment [19,20], management strategies [19-21], and settings [21]. The integrated ontology developed in this study is suitable for collecting social big data postings to answer various questions about the individual characteristics, management strategies, and environmental settings of obesity.

Our ontology incorporates terminology. The terminology of our ontology includes not only the terms extracted from the guidelines [22,23] but also various terms extracted from

social media postings. For example, there were ‘gynecoid obesity’ and ‘pear-shaped obesity’ as synonyms for ‘lower body obesity.’ The terminology with synonyms representing obesity-related concepts enables a rich collection of terms in social big data, and we utilized this feature in the keyword density has increased analysis on obesity types and management strategies.

Each class of our ontology has a data model with properties and possible value sets. The properties with values allowed us to conduct a sentiment analysis of obesity-related social big data. The results of the sentiment analysis showed that the rate of positive comments was higher than the rate of negative comments, and the rate of positive comments tend to increase annually during the analysis period. The postings related to effective exercise plans and diet were showed a positive awareness, while the postings related to the types of obesity and obesity as a risk factor for various chronic diseases were showed a negative sentiment on obesity and diet.

We evaluated the coverage, and structure and representative ability of our ontology. Our ontology reflected the concepts and terms that could appear in obesity-related postings from Twitter with a rather higher coverage rate above 87%. Our ontology was well represented in structure and representative ability with high scores above 4.0 out of 5 for all of the evaluation items. Some criteria, such as the connectivity between concepts and the relationships and accuracy of some superclasses were evaluated with somewhat low scores. We have modified and revised our ontology by following the opinions of evaluators.

When dealing with large data sets such as social big data, it is difficult to manage and manipulate data without a clear understanding of the concepts, structures, and characteristics of the data sets under investigation [1]. An obesity ontology as a shared conceptualization has potential uses in representing and managing obesity-related data [25]. An obesity ontology with data models and terminology of class concepts has even greater potential uses for representing and managing obesity-related unstructured social big data.

Since our ontology was developed for collecting and analyzing obesity-related social big data, it cannot provide answers to any biomedical questions on obesity, such as the underlying genetic mechanisms. It is quite possible that our ontology does not encompass all of the synonyms of the concepts because new terms will be continuously generated. Therefore, the terms representing the ontology concepts should be updated on a continuous basis.

Our findings suggest that the ontology developed in this study is suitable for collecting and analyzing social big data.

A keyword density analysis related to obesity types and management strategies, and a sentiment analysis of obesity and diet using social big-data postings were possible with our ontology. Our ontology can be used in future studies to answer competency questions not studied in this study.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This work was supported by a research grant (Efficient Management of Big Data on Health & Welfare) funded by the Korea Institute for Health and Social Affairs (No. 20140225), and by a National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2010-0028631).

References

1. Kuiler EW. From big data to knowledge: an ontological approach to big data analytics. *Rev Policy Res* 2014;31(4):311-8.
2. Song TM. Efficient utilization of big data on healthcare and welfare area. *Healthc Welf Forum* 2012;193:68-76.
3. Song TM, Ryu S. Big data analysis framework for healthcare and social sectors in Korea. *Healthc Inform Res* 2015;21(1):3-9.
4. Harris JK, Moreland-Russell S, Tabak RG, Ruhr LR, Maier RC. Communication about childhood obesity on Twitter. *Am J Public Health* 2014;104(7):e62-9.
5. Pagoto S, Schneider KL, Evans M, Waring ME, Appelhans B, Busch AM, et al. Tweeting it off: characteristics of adults who tweet about a weight loss attempt. *J Am Med Inform Assoc* 2014;21(6):1032-7.
6. Kuebler M, Yom-Tov E, Pelleg D, Puhl RM, Muennig P. When overweight is the normal weight: an examination of obesity using a social media internet database. *PLoS One* 2013;8(9):e73479.
7. Song TM, Song J, An JY, Hayman LL, Woo JM. Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends. *Yonsei Med J* 2014;55(1):254-63.
8. Alt R, Wittwer M. Towards an ontology-based approach for social media analysis. *Proceedings of the 22nd European Conference on Information Systems*; 2014 Jun

- 9-11; Tel Aviv, Israel. p. 1-10.
9. Yu EJ, Kim JC, Lee CY, Kim NG. Using ontologies for semantic text mining. *J Inf Syst* 2012;21(3):137-61.
 10. Gruber TR. A translation approach to portable ontology specifications. *Knowl Acquis* 1993;5(2):199-220.
 11. Noy NF, McGuinness DL. *Ontology development 101: a guide to creating your first ontology* [Internet]. Stanford (CA): Stanford University; c2015 [cited at 2017 Jul 20]. Available from: <http://www.lsi.upc.edu/~bejar/aia/aia-web/ontology101.pdf>.
 12. Hammond RA, Levine R. The economic impact of obesity in the United States. *Diabetes Metab Syndr Obes* 2010;3:285-95.
 13. Korea Centers for Disease Control & Prevention. *Korea health statistics: Korea National Health and Nutrition Examination Survey 2013 (KNHANES VI-1)* [Internet]. Cheongju: Korea Centers for Disease Control & Prevention; c2015 [cited at 2017 Jul 20]. Available from: https://knhanes.cdc.go.kr/knhanes/sub04/sub04_03.do?classType=7.
 14. Kang JH, Jeong BG, Cho YG, Song HR, Kim KA. Socioeconomic costs of overweight and obesity in Korean adults. *J Korean Med Sci* 2011;26(12):1533-40.
 15. Cristani M, Cuel R. A survey on ontology creation methodologies. *Int J Semant Web Inf Syst* 2005;1(2):49-69.
 16. Hammar K. *Towards an ontology design pattern quality model* [dissertation]. Linkoping, Sweden: Linkoping University; 2013.
 17. Scala PL, Di Pasquale D, Tresoldi D, Lafortuna CL, Rizzo G, Padula M. Ontology-supported clinical profiling for the evaluation of obesity and related comorbidities. *Stud Health Technol Inform* 2012;180:1025-9.
 18. Jeon BJ, Ko IY. Ontology-based semi-automatic construction of Bayesian network models for diagnosing diseases in e-health applications. *Proceedings of Frontiers in the Convergence of Bioscience and Information Technologies (FBIT)*; 2007 Oct 11-13; Jeju Island, Korea. p. 595-9.
 19. Zaragoza I, Alcaniz M, Botella C, Banos R, Guixeres J. An ontology for cognitive behavioral therapy: application to obesity. *Stud Health Technol Inform* 2009;144:16-8.
 20. Kim HY, Park HA, Min YH, Jeon E. Development of an obesity management ontology based on the nursing process for the mobile-device domain. *J Med Internet Res* 2013;15(6):e130.
 21. Shaban-Nejad A, Buckeridge DL, Dube L. COPE: childhood obesity prevention [knowledge] enterprise. *Proceedings of the 13th Conference on Artificial Intelligence in Medicine (AIME)*; 2011 Jul 2-6; Bled, Slovenia. p. 225-9.
 22. National Institute for Health and Clinical Excellence. *Obesity: the prevention, identification, assessment and management of overweight and obesity in adults and children*. London: National Institute for Health and Clinical Excellence; 2006.
 23. Kang JH, Kang JG, Kang JH, Kim KG, Kim D, Kim KS, et al. *Clinical practice guidelines for overweight and obesity in Korea*. Seoul: Korean Society for the Study of Obesity; 2012.
 24. Kehagias DD, Papadimitriou I, Hois J, Tzovaras D, Bateman J. A methodological approach for ontology evaluation and refinement. *Proceedings of the ASK-IT Final Conference*; 2008 Jun 26-27; Nuremburg, Germany. p. 1-13.
 25. Issa RR, Mutis I. *Ontology in the AEC industry: a decade of research and development in architecture, engineering, and construction*. Reston (VA): American Society of Civil Engineers; 2015.