Personality Measurement Design for Ontology Based Platform using Social Media Text

Andry Alamsyah*, Sri Widiyanesti, Rizqy Dwi Putra, Puspita Kencana Sari

School of Economics and Business, Telkom University, Bandung 40257, Indonesia

Abstract

Human behavior quantification is an essential part of psychological science. One of the cases is measuring human personality. Social media provide rich text, which can be beneficial as a data source to get valuable insight. Previous researches show that social media offered favorable circumstances for psychological researchers by tracking, analyzing, and predicting human character. In this research, we propose a personality measurement design to help to assess human character through linguistic usage from human digital traces. We construct our model by classifying social media text to the pre-determined personality facet from Big Five personality traits, mapping the knowledge to the ontology model, and implementing the model as a platform dictionary. Our model is based on the Indonesian language, which to the best of our knowledge is the first in the subject area. The platform is running effectively by using a well-established sorting algorithm, called the radix tree. Our objective is to support psychological science in adapting to a new technological era.

1. Introduction

The presence of advanced technologies, such as social media platforms and mobile devices, are shifting the way on how people communicate. Social media users have the freedom to upload daily routine information, exchanging messages, or even basic conversations [1]. The content created by the users is formed into digital traces. The personality of users revealed through their writing or textual content [2]. Each personality has its own charm that deserves attention over its complex arrangement. There are thoughts, hearts, and feelings that can change over time [3]. Hence, measuring human personality is a hard problem regarding the dynamic feeling's alteration.

Personality measurement has become the most extensive research in the field of Psychology [4]. The legacy methodology of personality assessment performed by interview and written examination [5]. Those methods are integrated method, where an interview is conducted to validate the test result. The characteristic of legacy methodology requires fulfillment instruments, such as the physical tools and psychologists. In order to adapt to the new technological era, some alternative methodologies are developed for bringing a faster process and result. One of them is the measurement through its natural environment or using digital traces on social media [6]. In this research, we utilize digital traces to put forward an automation personality measurement for minimizing fulfill instruments in the legacy methodology. Thus, it would significantly reduce the cost and reduce the time process of getting the results. This approach needs to be developed, considering that automation is the most demanded characteristic of the Industry 4.0 era.

According to Madden et al. [7]. digital traces, which consist of user information, e.g., personal information, shared texts, pictures, and videos, are a proof dataset that cannot be ignored, which expressed online human activity. These footprints are offered valuable opportunities for psychology research in understanding human characteristics [8]. Previous researches have assessed human personality through social media, such as Bhardwaj et al. [9], who assess personality through Facebook and LinkedIn and Park et al. [6] who applying the regression model to predict human character based on social media language.

Most of the research generally uses a machine learning approach, in contrast to our study, which utilizing the ontology approach. Machine learning provides us some leverage, such as the speed of analysis regarding large-scale data [10]. It is also able to predict personality on various forms of data like text, speech, and image [11]. However, the machine learning approach has some weaknesses in processing each meaning and intention of words due to language uniqueness [12]. Ontology afford us a better understanding of contextual knowledge [13]. There is an opportunity to use ontology as a basis for measuring human personality through the words on linguistic usage. Thus, in terms of measuring human nature through social media textual data, the
ontology model provides a more accurate result than a machine learning approach as long as the collection word type in the ontology model’s corpus is prosperous. The following approach scheme is shown in Fig. 1.

![Personality Traits Model Scheme between Ontology and Machine Learning Approach](image)

Fig. 1 shows the difference between ontology and machine learning model schemes. On the ontology model scheme, the personality trait model established by mapping words that reflect personality as a model’s forming component. Those collections of words then verified by experts and utilized as a dictionary for measuring personality. This model is a representation of an expert’s knowledge and adequate to map more than one personality trait in the complex sentence. Meanwhile, the personality trait model on the machine learning model scheme is a depiction of a machine’s algorithm according to the labeling process. The machine interprets the label on the training dataset as a rule for predicting personality. Therefore, the approach of the machine learning model scheme is entirely different than ontology. The ontology model scheme is only mapping the words with reference to an expert’s knowledge, while machine learning is predicting based on the learning result of the machine.

To the best of our knowledge, research of personality measurement has put forward in several personality traits taxonomies based on specified purposes, such as for assessing job placement and natural human emotion. In this research, we want to use the most general taxonomy of human’s essential persona. The Big Five Personality theory gets a consensus of a general taxonomy regarding human personality traits [14]. It also able to represent and simplify the diverse characteristics of a human’s personality [15]. The Big Five Personality Traits is built by examining several unique human attributes in their linguistic usage. This theory also offers prized terms called Revised Neuroticism-Extraversion-Openness Personality Inventory (NEO-PI-R) metric, which facilitate the exposure psychological characteristic of broad trait [16]. In this research, we use the Big Five Personality Traits Theory with the NEO-PI-R metric as a domain knowledge of our ontology model. The personality divided into five domains and further divided to thirty facet scales. Our research also applies the lexical hypothesis as a basis for analyzing textual content in social media. Based on research of Raad and Miamic [17], the lexical hypothesis is a process of understanding the meaning of textual data. The lexical hypothesis is qualified to map the personality traits through words in a language [18].

In terms of model development, the needs of an open-source platform are essential. The advantage of creating the platform are model crowdsourcing, public corpus enrichment, correction, and verification. Hence, the platform is significantly enhancing the model’s value over time. Nowadays, a platform with an ontology-based model for measuring human personality is still rare. In favor of getting a better model for measuring personality, we implement the model into a platform. This research aim is to show the way on how we develop a design in mapping human character by utilizing the ontology model. The model constructed by a collection of words that refer to personality in Bahasa. Our study mapped 2,331 instances in different facets and traits — those instances used as the corpus in our platform.

2. Literature Review

This chapter provides theoretical foundations related to the personality measurement proposed model based on the ontology approach. Literature review is sorted according to the data flow from social media data collection, personality definition, the personality traits, the ontology as the knowledge representation literature, and at last, is the radix tree algorithm used to parse and sort social media texts.

2.1. User-Generated Content

According to Naab and Sehl [19], user-generated content (UGC) has three criteria, there are 1). UGC is characterized by a rate of personal contribution, 2). UGC must be disclosed, and 3). UGC is built outside the sphere of occupation and professional routines. Besides, Wyrwoll stated that user-generated content is a content that is published online via various platforms by its user [20]. The users are not only the person but also the organizations. The substance classified as a UGC has similar characteristics, such as publicly accessible to other users, need a creative effort to create, and not a result of expertise routines and practices. UGC also defined in many forms, like blogs, posts, chat, podcasts, images, videos, tweets, and many other ways. In this modern era, UGC is a substantial information source for discovering knowledge from human digital activities [21]. Existing studies reveal that there is a high correlation between personality and personal inclination [22]. UGC indirectly shows beneficial information, like the user’s demand, lifestyle, and personality. Hence, we utilize UGC in social media to measure human personality.

2.2. Personality

Personality is defined as a set of a person’s characteristics, including acting, thinking, and feeling. Personality also correlates with emotions, values, attitudes, and talents [23, 24]. These attributes are establishing a unique persona of an individual and differentiate one person from another [25]. According to Steemler, a person’s personality closely related to language usage in speaking or writing [2]. Language is the most prevalent and reliable tool for people to convey their internal thoughts and emotions in giving comprehension to others [26, 27]. Hence, the linguistic usage of a person is an essential subject in the field of psychology and communication.

Previous research has examined human personality through linguistic usage. Howlader et al. predict Facebook user’s personality through status and linguistic features by applying
According to Costa and McCrae’s study [16], NEO-PI-R was advanced in samples of middle-age and older adults. The NEO-PI-R included scales to measure six conceptually derived facets in each OCEAN. The scales show generous internal consistency, temporal cohesion, convergent and discriminant validity against partners and peer ratings. Table 2 shows the six facets which explaining each of the factors in personality traits.

### 2.3. Big Five personality Traits

Big Five Personality Traits is a model that identifies five characteristics of personality. It is also recognized as the OCEAN model, which stands for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN) [34, 35]. This model is related to the lexical hypothesis and stated that language in daily interactions is a reflection of the most personality characteristics [36]. Thus, the Big Five Personality Traits predict and describe essential personality differences. In favor of getting a clearer explanation, Rossberger in [37] describes five traits below:

| Personality Traits | Facet | Description |
|--------------------|-------|-------------|
| Openness           | Aesthetics | Artistic interests; believe in the value of art/do not love poetry |
|                    | Fantasy   | Imagination; have an expressive imagination/seldom daydream |
|                    | Actions   | Adventurousness; prefer variety/dislike changes |
|                    | Ideas     | Intellect; admire convoluted problems/avert discussion |
|                    | Feelings  | Emotionality; encounter intense/rarely get emotional |
|                    | Values    | Liberalism; tend to elect for liberals/believe in one true divinity |
| Conscientiousness  | Competence| Self-efficacy; accomplish tasks successfully/misunderstand the situation |
|                    | Order     | Orderliness; prefer to order/abandon a mess |
|                    | Dutifulness| Follow the rules/break the rules |
|                    | Achievement-Striving | Work hard/just do the task |
|                    | Self-Discipline | Get task done right away/waste the time |
|                    | Deliberation | Cautiousness; avoid error carefully/charge into things directly |
| Extraversion       | Warmth    | Friendliness; make friends effortlessly/hard to make friends |
|                    | Gregariousness | Love large parties/prefer to be alone |
|                    | Assertiveness | Take action/wait for others to get the way |
|                    | Activity-Level | Regularly busy/love to take it easy |
|                    | Excitement-seeking | Love excitement/dislike loud music |
|                    | Positive Emotions | Cheerfulness; radiate joy/am seldom amused |
| Agreeableness      | Trust    | Trust others/distrust people |
|                    | Compliance | Morality; never deceive/use false praise |
|                    | Altruism  | Make people feel pleasant/despise on others |

Table 1: Big Five Personality Traits

| Big Five Personality Traits | Definition |
|-----------------------------|------------|
| Openness                    | or usually called openness to experience: extent to which individuals exhibit intellectual curiosity, self-awareness, and individualism/nonconformance. |
| Conscientiousness           | extent to which individuals value planning, acquire the tenacity quality, and achievement oriented. |
| Extraversion                | extent to which individuals involved with the external world, encounter enthusiasm and other positive emotions. |
| Agreeableness               | extent to which individuals’ value mutual effort and social harmony, modesty, dignity, and trustworthiness. |
| Neuroticism                 | extent to which individuals deal with negative feelings and their propensity to emotionally overreact. |

According to Costa and McCrae’s study [16], NEO-PI-R was advanced in samples of middle-age and older adults. The NEO-PI-R included scales to measure six conceptually derived facets in each OCEAN. The scales show generous internal consistency, temporal cohesion, convergent and discriminant validity against partners and peer ratings. Table 2 shows the six facets which explaining each of the factors in personality traits.

### 2.4. Ontology and Knowledge Representation

Ontology is a collection of concepts which able to model terms of vocabulary into a domain knowledge [38]. From the perspective of computational science, ontology is explained as a concept to model the system structure. For example, the relevant entities and relationships that exist from observations are useful for specific purposes [39].

www.astesj.com
| Personality Traits | Facet | Description |
|--------------------|-------|-------------|
|                    | Straightforward-ness | Cooperation; easy to fascinate/have a sharp tongue |
|                    | Modesty            | Dislike being center of attention/think highly of myself |
|                    | Tender-mindedness  | Sympathy; sympathize with the destitute/believe in people |
|                    | Anxiety            | Worry about things/relaxed most of the time |
|                    | Angry-Hostility    | Get enraged easily/rarely get annoyed |
|                    | Depression         | Often feel miserable/feel satisfying |
|                    | Self-Consciousness | Easily frightened/not being intimidated easily |
|                    | Impulsiveness      | Immoderation; often be affected by other/easily resist temptations |
|                    | Vulnerability      | Panic easily/remain calm under pressure |

Ontology associated with discovering and modeling reality under particular perspectives [40]. It focused on the structure and nature of an object [41]. Ontology also pointed out to a representational knowledge which indicates the type of class or entity associated with the relationship of the subtype [42].

Ontology generally has fundamental form components, i.e., class, instance, and relation [43]. Class is referring to a set of multiple instances, like words and phrases. An instance is a scope that considered in the ontology domain knowledge. A relation is defined as a relation among classes or instances [44]. This method is flexible, easy to modify, understood by humans and machines, and able to integrate with machine learning. Ontology has at least four evaluation methods [45]: 1) The Golden Standard; 2) The Application-based; 3) The Human Assessment; 4) The Data-Driven. Our study needs validation and evaluation for each instance in the model before implementing it on the platform.

Recent research has shown that ontology is able to represent knowledge for measuring personality. Some of the studies are that applying linguistic feature analysis and ontology model to measure human nature through social media data [46]. Another research computes human personality derivation in the modern physiognomy domain [47]. [48] also shows that ontology is a representative way to measure contextual knowledge, by building a complex domain of music and see how it relates to the domain of personality. A few research examples were clarifying how ontology might contribute to the analysis, comprehension, and research about human behavior and psychological research. Most of these existed researches are mainly focused in the English language, in contrast to our research which concerns in Bahasa.

2.5. Radix Tree

The dictionary which consists of strings data type is very-time consuming. A sorting algorithm is needed in order to get a more efficient process. Radix Tree is one of several algorithms that able to sort data in a database. Radix tree also beneficial for constructing associative arrays that expressed through the keyword in the form of strings. According to Mauro, radix tree is worked by labeling edges with a sequence of strings rather than characters, and constraining chains of nodes into a single element. Hence, radix tree is running more efficiently compared to a regular tree [49]. The instance of radix tree displayed in Fig. 2.

![Figure 2: The Instance of Radix Tree Algorithm](image)

3. Methodology

Our idea is to collect social media posts of people who considered influential in society. Their posts most likely to be responded by the public, which will generate conversations, and often set standard for informal language in Bahasa. In the model construction, we map words and phrases to a certain personality, which will validate by experts. There are several steps to construct the proposed model. Those steps are data collection, data preparation, model construction, model validation, platform construction, and conclusion. The research workflow is shown in Fig. 3. In contemplation of earning comprehensive understanding, we also illustrate the conception of our research methodology in Fig. 4.

![Figure 3: The Proposed Model Workflow](image)

![Figure 4: The Conception of Research Methodology](image)
3.1. Data Collection

Our research utilizes real-world conversations on Twitter social media. A recent study shows that Twitter provides a valuable chance to study human behavior in a natural environment [1]. In the data retrieval process, we use three samples i.e., the famous user in Twitter social media with specific criteria. Our samples criteria are:

1. Verified accounts or having tweets with more than 1000 tweets or 500,000 followers.
2. Shows the latest activity with different tweets.
3. Shows many interactions with other accounts.
4. Not a protected account.

Based on those criteria, we have collected 13,047 tweets from three selected users. We comprehend that despite having the same principles in data collection, this collection of famous people who have different character tendencies, and this is the reason why we conduct this research.

3.2. Preprocessing

The data preprocessing is required to clear irrelevant data, such as URLs, symbols, and other terms, which is not beneficial for this research [50]. This process objective is to get substantial information over the data [51]. According to Khadim, preprocessing is a process of improving data quality while reducing barriers that will occur in the classification process [52]. The preprocessing is divided into several steps shown in Table. 3.

| Steps | Name of Preprocessing Steps | Definition |
|-------|-----------------------------|------------|
| 1     | Case Folding                | Transforming all capital letters to lowercase letters. |
| 2     | URLs Removal                 | Eliminating web page links |
| 3     | Symbol and Number Removal    | Delete the symbol and number in the document |
| 4     | Tokenizing                   | Splitting the sentences into a token |
| 5     | Phrases Lookup               | Check if there are any phrases in the token |
| 6     | Synonym Recognition          | Determine the synonym of a word and replace it based on a dictionary |
| 7     | Word Generalization          | Replacing a word into a more general word in order to reduce data redundancy |

3.3. Ontology Model Construct

This model is established by classifying words in each tweet into thirty classes in the NEO-PI-R metrics and afterward generalize it into Big Five Personality traits. Those collections of words are defined as a corpus or dictionary of our model. In this study, we state the personality class in Big Five Personality Theory as a class, facet in NEO-PI-R metric as a sub-class, and each composed word into an instance. For better understanding, we show our ontology domain model hierarchy in Fig. 5.

The hierarchy of our ontology model is settled in the bottom-up paradigm since this model starting from specific to general or instances to class [53]. There are no properties needed in personality measurement ontology model. As an instance, from Fig. 5., the words terluka, melelahkan, penderitaan, which means hurting, exhausting, and suffering are classified into vulnerability in Neuroticism class. These words arranged to vulnerability facets because it indicates the feelings of flimsy due to the risk of harm from some experiences.

![Figure 5: The Hierarchy of Personality Measurement Ontology Model](image)

3.4. Model Validation

The validation process is essential to prevent and minimize any bias from the previous classification process. The human assessment approach is an assessment method with the help of domain experts [45]. For this reason, we assign psychologists to validate our classification result before deployment. In this process, the expert ensures that the classified word has entered the correct personality trait according to the Big Five theory and NEO-PI-R metric.

We conduct a human assessment because a human can represent language in terms of circumstances. The words in Bahasa ordinarily have various meanings depending on contextual purposes. For example, the words bisa can be represented to “able” or “can” or “poison from the snake.” Hence, we need the psychologist to check and validate the ambiguity of the words that frequently appeared in people's linguistic usage, especially in social media.

3.5. Platform Construction

Our platform is designed as an open-source platform to encourage the public or crowds to get involved in model development. People can openly provide some enrichment, correction, and verification of the model. Web-based platform is suitable for our research purposes due to its characteristics compared to other forms such as a mobile app or desktop app. The features of this platform are to measure personality based on the input. The common format of the collection of short text sentences and large-scale text is in comma-separated value (csv); thus, we use this .csv format file as one of our input files. At last, we visualize the result in a radar chart as the output. The platform’s workflow displayed in Fig. 6.

In this research, we utilize Personality Measurement Ontology (PMO) Platform which has been built in our prior research [51]. The platform is running by employing a radix tree algorithm as an effort to sort and parse the sentence, then find the semantic similarities between the input text and the corpus. The pseudocode presented in Fig. 7. The trait calculation is conducted
by finding similar keywords between processed textual data and the existing corpus in the ontology model. If matched, the data then considered representing one or more personality traits depending on the number of keywords that paired.

Figure 6: The Platform Workflow

The algorithm of our platform consists of several stages, i.e., convert the model to radix tree, saving database, input processing (sentence and csv format), and trait calculation. The following explanation of each stage are:

1. Algorithm 1: convert the model to radix tree. The established model is sorted using the radix tree algorithm. This algorithm composed of a) csv input file, which a text file contains of collections of keywords, b) a radix tree algorithm named ‘tree’; c) an algorithm that is occupied by the data in csv called ‘trie’.

2. Algorithm 2: saving database (keywords, traits, and facets). Generate a row for each keyword, facet, and traits into a ‘database’.

3. Algorithm 3: input processing (sentence). Parsing sentence by utilizing ‘trie’. The result of this process is called ‘set_of_words,’ which contains detected keywords.

4. Algorithm 4: input processing (csv). Parsing csv data into the database. The result is ‘set_of_words’ which loaded by detected keywords.

5. Algorithm 5: trait calculation. Calculate (‘set_of_word’) to get the desired form in the personality measurement results.

| Algorithm 1 (convert model to radix tree) | Algorithm 2 (saving database) |
|------------------------------------------|--------------------------------|
| **input** csv (keywords)                | **input** csv=[keyword, trait, facet] |
| **output** radix tree (keywords)         | **output** database=csv          |
| 1. csv = input (csv file)                | 2. data=[keyword, trait, facet]  |
| 2. tree = new radix tree                  | 3. for each row in csv do:       |
| 3. for each row in csv do:               |   data.keyCode=row.keyCode          |
| 4. trie <= dump (tree)                    |   data facult=row.facet          |
| 5. output (trie)                          |   database <- import (data)      |

| Algorithm 3 (input processing (sentence)) | Algorithm 4 (input processing (csv)) |
|------------------------------------------|-------------------------------------|
| **input** sentence                      | **input** csv=set_of_sentence       |
| **output** result=csv                   | **output** result=trait composition |
| 1. array = input (set_of_words)         | 2. csv = input (csv file)           |
| 2. trie <= load (database)              | 3. load radix tree                   |
| 3. for each word in array do:           | 4. for each row in csv do:          |
|   if word is in trie:                   |   for each word in row do:         |
|   set_of_words <- word                  |     if word in trie:                |
| 5. output (set_of_words)                |     set_of_words <- word            |

To facilitate user’s convenience, the platform provides a personality measurement process in two ways: 1) by entering textual data such as opinions, statements, or conversations; 2) by uploading the csv file, which consists of numerous textual data. The interface of our platform is shown in Fig. 8.

4. Result and Analysis

In order to examine our platform, we measure the personality of the three samples through the platform. We calculate a set of sample tweets in the form of a csv file. The result visualized into a radar chart for getting a better comparison of each personality trait. The measurement results displayed in Table 4.

From our measurement result, each sample tends to have one trait as a dominant personality trait. It is a good result because human always has at least one dominant persona over all of the existed character. It also proves that our platform is able to measure all of the traits in the Big Five Personality theory.

As shown in Table 4., our samples have a similar result but not entirely identical. The result also indicates that our samples have a high score in Extraversion personality traits. It means that our samples often show positive feelings, friendliness, and activity-level on their social media activity. For example, @shilicious, who regularly displays his activities, both significant and trivial matters in his daily routine.

The second-highest score of our sample’s measurement result is Agreeableness personality trait. This character is representing individuals who have value cooperation and social harmony with another person. @bepe20 is repeatedly showing his compliance and altruism. As a famous professional football player in Indonesia, @bepe20 has a tremendous number of fans. He
generally answers the message of his supporters on social media. He also does not hesitate to praise and welcome others who even he does not know. Thus, this persona makes @bepe20 also has a high score in Agreeableness besides the Extraversion personality trait.

| Account       | Result                        |
|---------------|-------------------------------|
| @jojosoherman | ![Diagram](image1.png)        |
| @bepe20       | ![Diagram](image2.png)        |
| @shitlicious  | ![Diagram](image3.png)        |

Even though the platform has proven capacity in measuring personality through linguistic usage, it still has some limitations. Our platform cannot measure complex phrases with different contextual meanings. For example, the sentence *kena gila*, which means he or she expresses a fantastic feeling, is measured into Extraversion and Neuroticism personality traits. The words *kena*, which means impressive, is reflecting Extraversion and *gila*, which means crazy or stupid is representing Neuroticism. That instance shows our platform’s limitation, which only running by mapping word by word. Hence, our platform still needs enrichment and development for measuring sentences with phrases that consist of a different word with different personalities.

5. Conclusion

In this research, we explore human’s linguistic usage in social media to build an ontology model for measuring human character. We have successfully implemented the ontology model to a web-based platform for measuring personality automatically. In this study, the ontology form component defined by assigning personality as a class, facet as a sub-class, and words that are referring personality as instances. The platform is generally running well but cannot handle the complex phrases.

Our model helps us to measure human personality in social media. This research significantly contributed to a psychology study, especially in Indonesia. Our model is adequate to support a psychologist to speed-up the personality measurement process. It also can be improved for reading complex phrases by generating another algorithm. Thus, our suggestion for future research is to employ another parsing algorithm such as n-gram for receiving better results. We also suggest enriching the platform’s corpus by engaging the words that are reflecting personality from different social media. Another recommendation is adding weight to each classified word. Since we only measure the frequency of words, and not considering the context of the tweets, adding weight to the indexed words is required to detect context on a sentence.

References

[1] A. Java, X. Song, T. Finin, and B. Tseng, "Why We Twitter: Understanding Microblogging Usage and Communities," 9th WEBKDD and 1st SNA-KDD Workshop, San Jose, California, USA, 2007.

[2] G. Stemmler, and J. Wacker, “Personality, Emotions and Individual Differences in Physiological Responses,” Journal of Biological Psychology, vol. 84, July 2010, pp. 541-551.

[3] S. I. Wahjono, “Perilaku Organisasi,” Graha Ilmu, 2010.

[4] L. Li, A. Li, B. Hao, Z. Guan, and T. Zhu, “Predicting Active Users’ Personality Based on Micro-Blogging Behaviors,” PLoS ONE 9(1): e84997, January 2014.

[5] J. L. Farr and N. T. Tippins, “Handbook of Employee Selection 2nd Edition,” Taylor & Francis Group, 2017.

[6] G. Park, H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, M. Kosinski, D. J. Stillwell, L. H. Ungar, and M. E. P. Seligman, “Automatic Personality Assessment Through Social Media Language,” Journal of Personality and Social Psychology, November 2014.

[7] M. Madden, S. S. Fox, A. Smith, and J. Vitak, “Digital Footprints”, Internet: https://www.pewresearch.org/internet/2007/12/16/digital-footprints/, 2007 [October 18, 2019].

[8] W. Bleidorn, C. J. Hopwood, A. G. C. Wright, “Using Big Data to Advance Personality Theory,” Current Opinion in Behavioral Sciences, vol. 18, December 2017, pp. 79-82.

[9] S. Bhardwaj, P. K. Attrey, M. K. Saini, A. E. Saddik, “Personality Assessment using Multiple Online Social Networks,” Multimedia Tools and Applications, vol: 75November 2016, pp. 13237-13269.

[10] B. Y. Pratama and R. Sarno, “Personality Classification Based on Twitter Text Using Naive Bayes, KNN, Aand And SVM,” 2015 International Conference on Data and Software Engineering (ICoDSE), November 2015.

[11] G. Farnadi, G. Sitaraman, S. Sushmita, F. Celli, M. Kosinski, D. Stillwell, S. Davalos, M. F. Moens, and M. D. Cock, “Computational Personal Recognition in Social Media,” User Modeling and User-Adapted Interaction, vol:26, June 2016, pp. 109-142.

[12] B. Liu, L. Yao, and D. Han, “Harnessing Ontology and Machine Learning for RSO Classification,” Springer, 2016.

[13] N.F. Noy, and D.L. McGuinness, “Ontology Development 101: A Guide to Creating Your First Ontology,” Stanford Knowledge Systems Laboratory, Technical Report KSL-01-05, 2001.

[14] O. P. John, R. Robins, and L. Pervin, “Handbook of Personality Theory and Research,” New York: The Guilford Press, 2008.

[15] P. T. Costa and R. R. McCrae, “Normal Personality Inventory,” Psychological Assessment, 4(1): 5-13, March 1992.

[16] P.T. Costa Jr and R.R. McCrae, "Domains And Facets: Hierarchical Personality Assessments using The Revised Neo Personality Inventory.,” Journal of Personality Assessment, vol. 64, February 1995, pp. 21-50.

[17] B. D. Raad and B. Mlaciec, “The Lexical Foundation of The Big Five Factor Model,” in oxford handbook of the Five Factor Model, US: Oxford University Press, 2017, pp. 191-216.

[18] B. De Raad, "The Big Five Personality Factors: The Psycholexical Approach to Personality,” Hogrefe & Huber Publishers, Göttingen, 2000.

[19] P. K. Naab and A. Sehl, “Studies of User-Generated Content: A Systematic Review,” Journalism 18(10), Sagepub, October 2016.

[20] C. Wyrrwoll, “Social Media Fundamentals, Models, and Ranking of User-Generated Content,” Wiesbaden: Springer Fachmedien Wiesbaden, 2014
