March - April 2020
ISSN: 0193-4120 Page No. 8150 - 8158

A Proposed Concept and Platform for Ontology based Personality Measurement Model

Andry Alamsyah1, Muhammad Fadhli Rachman2, Cindy Septiani Hudaya3, Rimba Pratama Putra4, Aulia Ichsan Rifkyano5, Fivi Nurwianti6

[1,2,3,4] School of Economics and Business, Universitas Telkom
[5] School of Electrical Engineering and Informatics, InstitutTeknologi Bandung
[6] Faculty of Psychology, Universitas Indonesia
[1]andrya@telkomuniversity.ac.id ,[2]fadhlirachman98@gmail.com , [3]cindyshudaya@gmail.com , [4]pratamarimba@gmail.com , [5]rifkyanoa@gmail.com , [6]fivinurwianti201@gmail.com

Article Info
Volume 83
Page Number: 8150 - 8158
Publication Issue:
March - April 2020

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 09 April 2020*

Abstract

Personality measurement (PM) has evolved from taking a personality test or expert assessment into utilizing user generated content to measure personality. Several attempts to create a model for PM has been done by the various researcher using English as the main language. While the PM model in Bahasa Indonesia is limited. Further research is required due to the unavailability of PM model in Bahasa Indonesia, lack of corpus to gain high accuracy, and the urgency of automating the measurement process. In this paper, we develop the PM ontology model by 1) enriching the corpus in Bahasa Indonesia 2) design and prototyping the PM ontology platform 3) evaluate the PM ontology model. Our proposed platform offers a fast and affordable tool to analyze large textual data to measures human personality based on big five personality traits. Extensively, the platform is beneficial to have express analysis process and utilized the insight into various areas such as human resources, CRM, marketing or another process that requires personality measurement.

Keywords; *Personality Measurement, Ontology Platform, Ontology, Big Five Personality Traits*

I. INTRODUCTION

The personality test is becoming one of the determinant factors on human resources management area. Personality test act as a complementary factor to cognitive test in the selection process to get high performers [1]. Its implementation usually brings issues, one of them is that candidates tend to respond favorably and even faking the responses [2]. To avoid the made-up responds, learning personality through from its behavioral residue in a natural and virtual environment such as digital trace from social media is recommended [3].

A model to measure personality based on User Generated Content (UGC) posted on social media already exists. The model itself, using big five personality traits and NEO-PI-R as its basic theory to measure the personality [4]. Measuring personality using UGC has been done before using Naïve Bayes, this machine learning method has weaknesses in processing large textual data with high level of noise and incompleteness [5]. Ontology modeling approach is able to overcome the machine learning weaknesses and has been used to a various area such as academic evaluation system, classify documents [6], and wine classification [7] Ontology approach is also used to map the knowledge of Bahasa Indonesia words to represents each facet and traits in NEO-PI-R [6]. In

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 8150 - 8158

our previous research, the PM model has mapped 343 words and idiom into facets and traits.

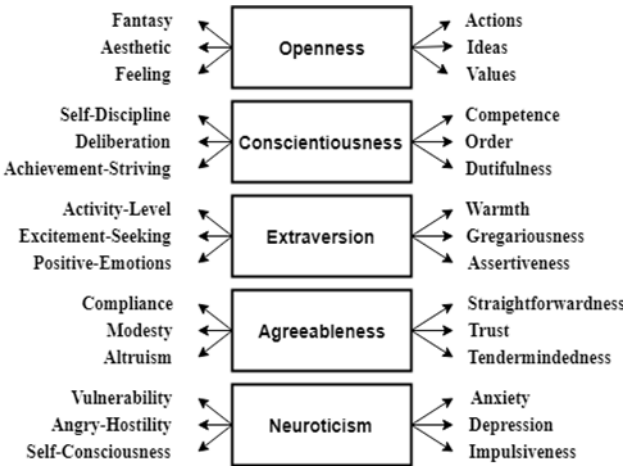
Ontology evaluation are required to ensure the ontology model is good to be used by human and machine. There are four methods to evaluate ontologies [8] such as 1) golden standard which comparing the ontology to established model; 2) application based evaluation by measuring the ontology performance to a standard ; 3) human assessment which using domain experts to do the evaluation; 4) data-driven evaluation that measures the quality of ontology using formula. Another way to measure the ontology model represents knowledge or not is by measuring the semantic variance [9]. But the semantic variance will not be discussed in this paper due to the lack of ontologies that covering the same domain knowledge. The semantic variance requires to select the most accurate model from a collection of the same ontologies [9]. To decide the types of ontologies we build, we used of kind of ontologies which categorize ontology model to catalog, glossary, thesaurus, taxonomy, and proper ontology [10].

We enrich the corpus by 1988 instances that represent facets that still following the previous model constraints such as Bahasa Indonesia and NEO-PI-R. We also create a PM ontology platform to facilitate the crowd to access and contribute to the PM ontology model easily. Lastly, we discuss how this model is evaluated.

The platform is able to process automatically and analyze large- scale textual data such as opinions, statements, and conversations. The platform itself using five algorithm such as converts model to Radix tree, saving database, input processing sentence and csv, and trait calculation. Therefore, this research has the capacity to help a lot of areas and not limited to human resources, marketing, or other selection process that requires personality measurement.

II. LITERATURE REVIEW AND RELATED WORKS

A. Big Five Personality



To identify relevant studies on the relationships between Big Five Personality traits and ontology, we followed the literature search as proposed in [6]. Digital footprints can predict personality traits and can be used as faster survey alternatives, lower costs and support a much larger human population [11]. We conducted the development of the existing ontology model to enrich the vocabulary. We use the same metric in this research which is shown in Figure 1.

B. Personality Measurement Platform

The ontology development in this research demands the creation of platforms such as web applications with large personality database models used to measure human personality. Web applications are chosen because they are easy to manage and do not require high maintenance costs. Implementation of the ontology of the web has been done a lot, it can be seen from the previous study such as in [12] Lei et al. (2005) talk about a new software tool Onto Weaver, that uses an ontology to drive the design and development of their data-intensive website. While in [13] López-Cima et al. (2007) discuss the development of on-line web applications using the JSTL expression language through components, which are used for developing portals that focus on dissemination and management of their research

Published by: The Mattingley Publishing Co., Inc.

Fig. 1. *NEO-PI-R* metric

R&D development collaboration projects. Then in [6] Alamsyah et al. (2018) they suggest that further research can implement automated machine learning software, in order to more easily integrated models, and also can facilitate the application of existing ontology model. Therefore, we applied that suggestion in this study by creating a platform that uses the radix tree algorithm to optimize the process on that platform. The architecture of the PM Ontology platform can be seen in the Figure 4.

C.Building Ontology Model

The construction of the ontology model with the big five personality was carried out previously in [6] Alamsyah et al. (2018) they enriched the size of the human personality from the writing and language observations used on social media with the aim of completing personality tests that previously only used survey tests or questionnaires. The ontology model in their research is used to classify the text of the construction the human personality in Indonesian. However, they did not take measurements using the web application platform as we did in this study. Besides that, the number of words we used for personality measurements in this study were also more than in previous studies. For more details can be seen in Table III.

D. Ontology Model Evaluation

There are many definitions of what an ontology is. One of the common definitions of ontology is the domain knowledge concept that can be understood by machines [6]. Ontology as a knowledge representation has a role in formalizing knowledge so that ontology can be replicated or automated. If we used it properly, ontology it is a powerful tool. With ontology, the computer can operate based on formal semantic knowledge representation and allows people to understand the processes in the knowledge base. The knowledge base is divided between several types of automatic and non-automatic reasoning that are plotted against the knowledge base. This research uses automatic reasoning called Taxonomy [8].

As discussed previously, ontology evaluation has several approaches, namely the golden standard, application-based, human assessment, and data-driven. In this section, we discuss deeper about these approaches. Golden standard evaluation makes it possible to conduct multidimensional ontology evaluations in an automated manner. In the golden standard, ontologies are evaluated to determine the optimal level of the ontology in the domain based on certain standards set previously. However, to evaluate the golden standard approach, we must have the gold standard ontology first. The application-based approach is considered quite promising because the complexity of the method in evaluating this ontology can be reduced according to the complexity of the specified task evaluation. However, this approach also has the disadvantage of only testing one possibility from the ontology aspect, which can cause the measurement of the overall error. The Human assessment approach is the only assessment method that can be assessed for each criterion from the specified list. This assessment can be done with the help of experts who can pay attention to inconsistencies and incorrect or missing information. The data-driven approach can assess ontology in a way that humans do, by extracting domain knowledge from data. However, this method takes into account only one criterion with a certain data-driven approach and of course limited [8].

From several methods to evaluate ontologies, we use the human assessment approach as an evaluation ontology in this study. Human assessment of ontologies can be divided into two main categories: the qualitative and quantitative human assessment. [8] We use qualitative evaluation in this study to measure results based on the relationship between the Big five personality and Ontology. Human assessment as an evaluation method it works well because humans can represent language in terms of

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 8150 - 8158

reasoning and this ontology evaluation method can also provide scores on predetermined criteria [8]. For the ontology model construction paradigm, in this study, we use the bottom-up paradigm because the paradigm can define specific entity models to common model entities, which are in accordance with the needs of this research [6].

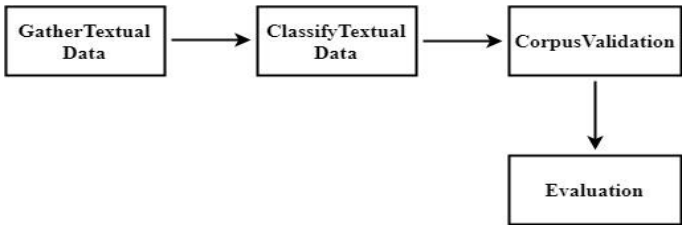


Fig.2. Ontology Evaluation Workflow

The workflow stage in the ontology evaluation begins with collecting data in the form of text which is then classified to proceed to the validation stage. After the validation process was completed, we proceeded to conduct an ontology evaluation.

III. METHODOLOGY

This model classifies social media textual data into any big five personality traits [6]. After we evaluate the existing model, we build the PM ontology model into a web-based platform that can analyze textual data to measure personality automatically. The entire experimental process is shown in Figure 3.

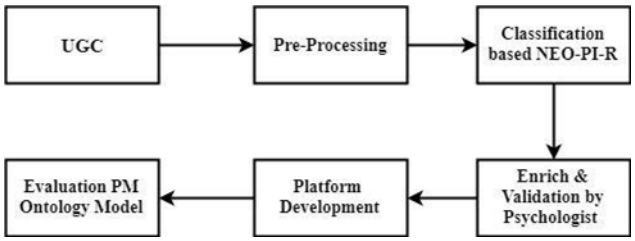


Figure 3

A. Twitter Data

We utilize Twitter conversational data from 3 samples. The samples textual tweets gathered to enrich the instances of the PM ontology model. Our sample requirements are: 1) public figure accounts that are verified accounts and not protected accounts, 2) actively discuss or interact with other users, give lots of opinions, or share many things about their daily lives. By using Twitter API, we get 9602 tweets from 3 accounts and reduce the tweets to 6398 by removing retweets data. The samples from our research are explained below in Table I.

Table I Data Samples Collection

Twitter Samples	Tweets Crawled	Without Retweet
@jojosuherman	3173	1773
@bepe20	3235	1835
@shitlicious	3194	2780
Total	9602	6398

B. Pre-Processing

We preprocess data samples in order to remove retweets, remove URLs, symbol, and other terms which are not assigned to any of the big five personality traits to produce more meaningful information. The preprocessing steps are shown below in Table II.

Table II Pre-Processing Steps

No	Pre-Processing Steps	Description
1.	Retweets Removal	Remove retweets that exist on the document
2.	Username and URL Removal	Remove links and username on a tweet.
3.	Non-Alphanumeric Removal	Remove number and symbols on a tweet and textual data
4.	Keyword or World Finding	Find the word that can reflect personality traits
5.	Synonym Detection and Word Generalization	Find same meaning words that exist on the textual data and substitute them with to suited general.
6.	Idiom Lookup	Find idioms on the textual data and considered them into one term.

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 8150 - 8158

C. Classification based NEO-PI-R

Classification based NEO-PI-R is classifying each tweet into 30 classes based on big five personality traits and its facets. The 30 classes are generalized into five personality traits. In the classification process, the authors find keywords on each preprocessed tweet manually. Those collected keywords are inserted into a class to represent a personality trait. From preprocessed data, we collect as many as 1988 words that can reflect personality traits as a component of our model. We classified all words into special classes based on the big five personalities that produced the corpus.

D. Validation by Expert

Given that the classification process is done manually and has a strong dependency on the author's perception of the keywords, the classification has a high possibility of bias. To prevent and minimize the bias, the corpus will be tested and validated by expert psychologists. PM ontology model for measuring human personality is based on the validation results.

E. Building Personality Measurement Ontology Model

We build PM ontology model by instances which are a collection of keywords from social media that have been collected to determine a personality type. In the previous stage of research, this has been done by classifying tweets to find keywords on social media that can reflect personality. After the building phase is done, the ontology model is stored in the Ontology Web Language (OWL) file. The owl document can not only explain the ontology data model, but also the object word that describes its property and class, relations between classes, hierarchy, properties, and characteristics of its properties.

F. PM Ontology Platform

We build a platform consists of the PM ontology model and algorithms to measure the personality. Admin build first generation of the PM ontology model. The model will be converted into Radix tree which finds semantic similarities between the text in the document and the corpus from persistent storage using the Radix tree. The architecture of the PM Ontology Platform is shown in the Figure 4.

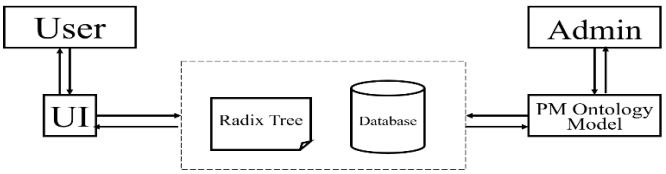


Fig. 4. The architecture of the PM Ontology Platform

Furthermore, we use the Radix tree to optimize search and space in the database [14]. So that the process is more efficient. The algorithm of the PM ontology platform consists of convert model to radix tree, saving database, input processing sentence, input processing CSV, and trait calculation. The following is an explanation of each of the stages:

1. Convert model to Radix tree: The model created is classified using the radix tree algorithm. In this algorithm consist of 1) ‘csv’ which a file consists of keywords 2) a radix tree algorithm named ‘tree’ 3) an algorithm that is filled by the data in csv named ‘trie’.

2. Saving Database (keywords, traits, and facets): In this process we create a row for each keyword, facet, and traits. Then, import them to ‘database’.

3. Input processing (Sentence): Parsing sentence using ‘trie’, and the result is 'set_of_words' which contains detected keywords.

4. Input processing (CSV): Parsing CSV data into the database table and the result is 'set_of_word' which contains detected keywords.

5. Trait Calculation: The results of the previous input process (‘set_of_word’) are calculated to get the desired results in the form of traits.

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 8150 - 8158

Algorithm 1	**Convert Model to Radix tree**
Input	CSV (keywords)
Output	Radix Tree (Keywords)
1	`csv = input (csv file)`
2	`tree = new Radix tree`
3	**`for each`** `row in csv`**`do:`** `tree ← row. keyword`
4	`trie ← dump (tree)`
5	**`output`** `(trie)`
Algorithm 2	**Saving Database (keywords, traits, and facets)**
Input	CSV (keywords, traits, and facets)
Output	Database (keywords, traits, and facets)
1	`csv = input (csv file)`
2	`data = [keywords,trait,facet]`

3	**`for each`** `row in csv`**`do:`**
	`data.keyword = row.keyword`
	`data.trait= row.trait`
	`data.facet= row.facet`

4	`database ← import(data)`
5	**`output`**`(database)`
Algorithm 3	**Input Processing (Sentence)**
Input	Sentence
Output	Result (set of words)
1	`array = input(sentence)`
2	`trie ← load (database)`
3	**`for each`** `word in array` **`do:`** **`if`** `word is in trie:` `set_of_words← word`
4	**`Output`** `(set_of_words)`
Algorithm 4	**Input Processing (CSV)**
Input	CSV (set of sentences)
Output	Result (set of words)
1	`csv = input (csv file)`
2	`load Radix tree`
3	`trie ← load (database)`
4	**`for each`** `row in csv`**`do:`** **`for each`** `word in row` **`do:`** **`if`** `word is in trie:` `set_of_words ← word`
5	**`output`** `(set_of_words)`
Algorithm 5	**Trait Calculation**
Input	Results (set of words)
Output	Result (traits composition)
1	`array = input set_of_words)`

2	**`for each`** `word in array:`
	`trait = trait (where`
	`facet = word.facet)`
	`trait =+ 1`

3	**`output`** `(trait)`

Fig. 5. The Algorithm of the PM Ontology Platform

To simplify and reduce the time of measuring personality, we offer a ubiquitous platform to measure and visualize textual data. The user is able to input texts or CSV file in the PM ontology platform. Then, the output is a radar chart that shows the percentages of each trait. The measurement process on the PM ontology platform is shown in Figure 6. We also compared training data to evaluate the measurement results from the PM ontology platform with the help of a psychologist for the level of accuracy.

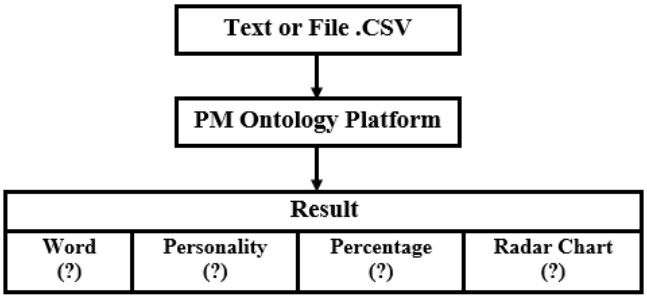


Fig. 6. The Measurement Process on the PM Ontology Platform

IV. RESULTS AND DISCUSSION

The corpus increases as much as 2331 instances to develop the PM ontology model shown in Table III. From 1988 new instances it is consist of 543 new instances for Agreeableness, 388 new instances of Neuroticism, 385 new instances of Openness, 338 new instances for Conscientiousness, and 334 new instances for Extraversion. These new instances will contribute to map more Bahasa Indonesia words in measuring personalities.

Table III. Instances On Ontology Model

	Tweets	Instances
Previous Model	3445	343
Updated Model	9602	1988

Figure 7 is the interface of the PM ontology platform, that crowd can take measurements in 2 ways i.e. 1) by entering textual data such as opinions, statements, or conversations. and 2) by uploading the CSV file which consists of numerous textual data.

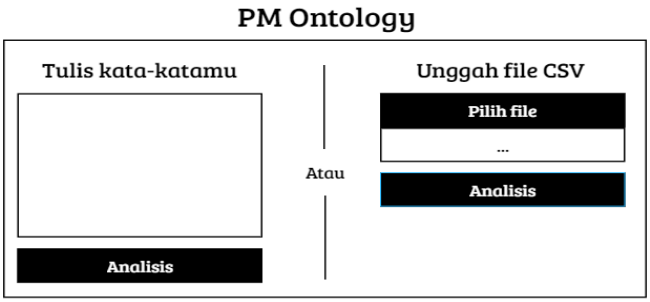
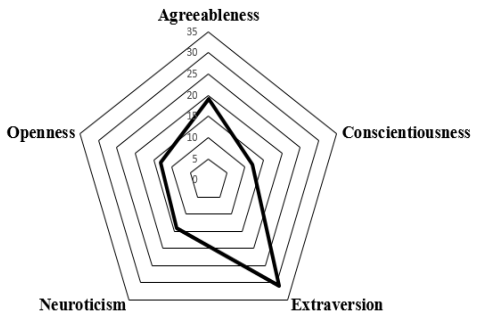


Fig. 7. The Interface of the PM Ontology Platform

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 8150 - 8158

We conduct evaluations using two measurement instruments, namely PM ontology and expert evaluation. Both of these measurements are using the same data set but has different results. The difference in results obtained is not surprising. Because, the focus of two measurements instruments are also different. PM Ontology measurement result is focuses on word frequency its detected; it can be seen in Figure 8. While the expert evaluation result using the same dataset focuses on the context of the conversation, the result shown in Figure 9 and Table IV. Therefore, an important evaluation for this ontology model is to develop PM ontology measurements that do not only focus on detecting the frequency of terms but can also detect the context of the textual data.



Result

Total words : 99

Agreeableness: 16 (16.16%)

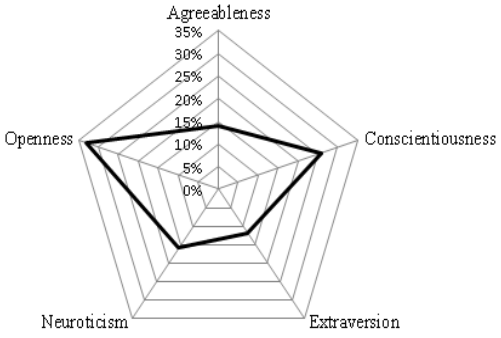
Conscientiousness: 10 (10.1%)

Extraversion: 31 (31.31%)

Neuroticism: 14 (14.14%)

Openness: 13 (13.13%)

Fig. 8. The Result Page of the PM ontology platform



Result

Total words : 41

Agreeableness: 6 (14%)

Conscientiousness: 11 (26%)

Extraversion: 5 (12%)

Neuroticism: 7 (16%)

Openness: 14 (33%)

Fig 9. The Psychologist Result

Table IV Sample Classification Result In Context

Keyword	Facet	Traits
ga_setuju	Deliberation, Straightforwardness	Conscientiousness, Agreeableness
sama	Deliberation, Straightforwardness	Conscientiousness, Agreeableness
saya_kira	Self-Consciousness, Deliberation	Neuroticism, Conscientiousness
saya_yakin	Self-Consciousness, Competence	Neuroticism, Conscientiousness
aduh_kenapa	Ideas, Anxiety	Openness, Neuroticism
jangan_ulangi	Assertiveness, Order	Extraversion, Conscientiousness

In addition, the crowd can contribute to enriching the ontology model by adding instances in the corpus. On the user panel page, the crowd able to enter instances independently, the instances are stored into a database on the PM ontology platform. Although, the expert validation process is still needed to confirm the instances has the right facets. The user panel is shown in Figure 9.

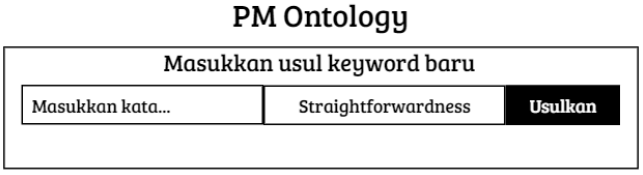


Fig. 10. The User Panel of the PM ontology platform

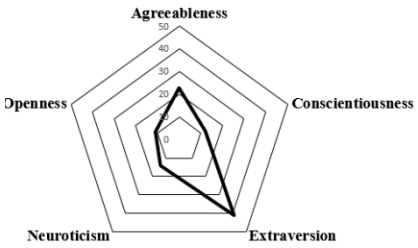
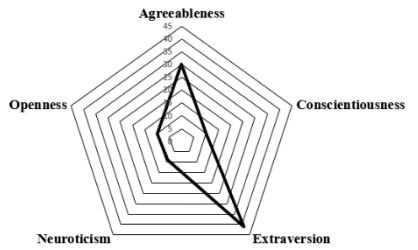
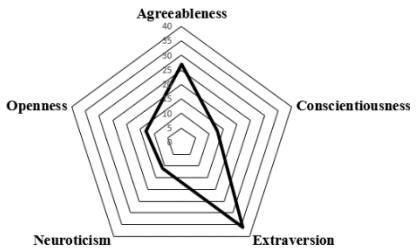
As an example, we calculate a collection of tweets in form of CSV file from twitter data account with

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 8150 - 8158

the PM ontology platform and the results are in the form of a radar chart that illustrates the big five personality traits. The measurement results are shown below in Table IV.

Table V The Measurement Result

Account	Result
@jojosuherman	
@bepe20	
@shitlicious	

The results of the PM ontology platform, that 3 sample accounts have a tendency of more than 1 personality trait but still have the main personality traits. This result can be increased by using richer data for its ontology model. In our sample, @jojosuherman, @bepe20 and @shitlicious have high extraversion, also have high agreeableness.

V. CONCLUSION

Our PM ontology model has richer instances than previous research to represent Bahasa Indonesia terms. We also create an automation on the personality measurement process and evaluated by an expert. Our suggestion for the next development is to keep enrich the model's instances and equalize the number of instances in each trait to prevent one trait overpowering other traits, we also suggest to implement low level and high-level instances to gain better accuracy in the measurement, building a formal ontology to represent general logic. Another suggestion is to determine the same focus of measurement between evaluations from psychologists and PM ontology models.

We conclude that the result from our platform is resulting in other questions and hypotheses which are 1) are verified account is providing us normative tweets? 2) is the number of instances in each trait should be equal?

REFERENCES

[1]. C. Kelly, "Personality Tests: A tool for Predicting High Performing Employees," November 2010. [Online]. Available: https://www.cpshr.us/documents/whitepapers/PersonalityTests.pdf. [Accessed 9 December 2018].

[2]. W. Arthur, D. J. Woehr, and W. G. Graziano, "Personality testing in employment settings: Problems and issues in the application of typical selection practices," Emerald Insight, vol. 30, no. 6, pp. 657-656, 2001.

[3]. L. Qiu, H. Lin, J. Ramsay, and F. Yang, "You are what you tweet: Personality expression and perception on Twitter," Journal of Research in Personality, vol. 46, pp. 710-718, 2012.

[4]. B. Liu, L. Yao and D. & Han, "Harnessing Ontology and Machine Learning for RSO Classification," Springer, 2016.

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

[6]. A. Alamsyah, M. R. D. Putra and D. D. Fadhilah, "Ontology Modelling Approach for Personality Measurement based on Social Media Activity," in ICoICT, Bandung, 2018.

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 8150 - 8158

[7]. N. F.Noy and D. L. McGuinness, Ontology development 101: a guide to creating your first ontology, Stanford: Research Gate, 2000, p. 3.

[8]. K. Sekandar, A Quality Measure for Automatic Ontology Evaluation and Improvement, Utrecht: Utrecht University, 2018.

[9]. D. Sanchez, M. Batet, S. Martinez and J. Domingo-Ferrer, "Semantic Variance: An Intuitive Measure for Ontology Accuracy Evaluation," Engineering Applications of Artificial, vol. 39, pp. 89-99, 2015.

[10]. M. Uschold and M. Gruninger, "Ontologies and Semantics for Seamlessly Connectivity," ACM SIGMOID Record, p. 2, 2004.

[11]. D. Azucar, D. Marengo and M. Settanni, "Predicting the Big 5 personality traits from digital footprints on social media: A meta-Analysis," Italy, 2018.

[12]. Y. Lei, e. Motta and J. Domingue, OntoWeaver: an ontology-based approach to the design of data-intensive web sites, UK: Research Gate, 2005.

[13]. A. López-Cima, O. Corcho and A. Gómez-Pér, "Rapid ontology-based Web application development with JSTL," in Research Gate, Madrid, 2007.

[14]. H. Zhang and G. Chen, "In-Memory Big Data Management and Processing: A Survey," IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 7, pp. 1920-1940, 2015.

Published by: The Mattingley Publishing Co., Inc.