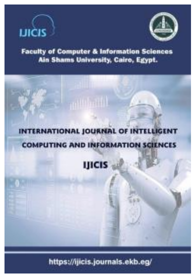
IJICIS, Vol.23, No.1, 18-28
DOI: 10.21608/ijicis.2023.137912.1181

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/



ARCHITECTURE FOR PERSONALITY DETECTION USING ENNEAGRAM KNOWLEDGE: CASE STUDY

Esraa A. Abdelhamid
Computer Science,
Faculty of Computer and Information Science, Ainshams University,
Cairo, Egypt
esraa_cs@hotmail.com

Sally S. Ismail
Computer Science,
Faculty of Computer and Information Science, Ainshams University,
Cairo, Egypt
sallysaad@cis.asu.edu.eg

Mostafa M. Aref
Computer Science,
Faculty of Computer and Information Science, Ainshams University,
Cairo, Egypt
mostafa.m.aref@gmail.com

Received 2022-05-11; Revised 2023-01-25; Accepted 2023-01-27

Abstract: *Researchers are concerned with automated personality detection from social media. Automated Personality detection from text benefits in social media like: attracting more users, career advising and getting more advertisements. Traditional personality detection is done by using an assessment test. Performing a test is time-consuming so users aren't interested in taking a test. This paper presents case study and architecture on automated personality from text using twitter text. The case study uses public text to identify the personality of the profile. The applied personality Model is the Enneagram model. Proposed architecture contains four phases: text preprocessing, feature extraction, feature selection and personality detection. Feature selection is done by using Enneagram ontology and lexicon. Personality detection is utilized by using statistical approaches. The Enneagram knowledge is modeled using ontology. The lexicon is a source to enrich the ontology seed. Statistical Approach is utilized to identify the personality. The case study identifies the personality. The highest outcome percentage is "investigator" personality, which is 24 %. This indicates that the personality is an investigator. This result is similar to official Enneagram experts' analysis. This model is the first one which uses the Enneagram model as automatic detection. Enneagram is a powerful personality model that aids psychiatrists and physicians to understand the patient's personality intensely. This knowledge gives them the tools to support and aid the patient to heal faster. The promising outcomes open the door to further research in this area.*

Keywords: *Data Mining, Text Mining, Personality Detection, Enneagram, Psychology.*

***Corresponding Author:** Esraa A. Abdelhamid

Computer Science, Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt

Email address: esraa_cs@hotmail.com

1. Introduction

Personality defines a lot about an individual. Knowing the personality of an individual is a gate to multiple benefits. Knowing people's behavior and preferences, helps us to significantly build better recommendation systems [1]. In social media, Personality detection is useful for attracting more users, increasing advertisement and career advising. The rise in author profiling is increased largely because of the massive volume of textual data produced by users [2]. Personality models can also help in online dating, education development and personal development. Researchers are concerned in many domains with personality detection like psychology, artificial intelligence, natural language processing, behavioral analytics and machine learning [3].

Personality assessment is the traditional way to measure personality. Persons are not eager to perform a questionnaire as it is not practical and time wasting [4]. Many users are not interested in doing a test. When using a personality test, one of the most common concerns is that respondents tend to answer according to favor, even faking the answers [5]. This tends to produce unrealistic results. Personality detection from text does not consume time. Personality detection catches attention in various societies like social media, robotics, human computer interaction and speech processing [6]. Personality detection is a hard and complex task for humans as it requires expert knowledge. Personality detection from text has become a hot topic for research. Personality traits can be predicted via digital footprints, which can also be utilized as a faster survey alternative with cheaper costs and support for a much bigger population [7]. Social Media provides a lot of public text which can aid in this area. Personal information from social media users can be utilized for a variety of purposes, including crowd behavior analysis [8].

There are a lot of personality models. These models are the big five personality model, Myers Briggs Type Indicator (MBTI), three factor model and Enneagram. Most of the past research focus on big five, MBTI and three factor models. These models measure certain traits in the personality without giving insights about the motivation of behavior.

Enneagram is a personality model which consists of nine personalities. Enneagram is mapping the nine basic personality types of human nature and its compound mutual relations [9]. Enneagram demonstrates the person’s fears, desires, motivation, stress and growth. A person can learn from the Enneagram his point of strengths, his limitations and targeting their growth [10]. Enneagram demonstrates human behavior in depth. Enneagram is studied at many USA universities in medicine, psychology, education, arts and business [11]. Enneagram also has other usages. Enneagram is a significant tool to enhance relationships with family, friends and coworkers [12]. Enneagram also assists counseling. Enneagram helps the counselor to detect client behavior, this knowledge increases healing and growth [13]. Enneagram also aids physicians and psychiatrists to understand the patient's personality. Recognizing the patient's personality gives physicians insight to provide the patient with psychological support. Psychiatrists have used the Enneagram since 1970 [14].

Personality identification using social network analysis is a relatively new domain within machine learning research [15]. Machine learning has several problems in dependability, transparency and consistency. In the context of natural language processing (NLP), these issues are particularly crucial because, unlike in other fields, they prevent AI from achieving human-like performance [16]. Machine learning techniques do not provide satisfying results as this task is more complex thus the need for sentiment analysis is aroused. In this research, Personality Detection from Text using sentiment Analysis is applied in this model.

The proposed architecture are text prepossessing, word-based feature extraction, word-based feature selection and personality detection. Text prepossessing is used to cleanse the text from unnecessary extensions and non-useful text. Word-based Feature Extraction is used to take out features from the text as a starting phase. Word-based Feature Selection is applied to choose important features as a filtering phase. This feature selection employs both ontology and English lexicon. Ontology provides a good knowledge about a specific domain [17]. Ontology was utilized as a knowledge representation technique. Ontology is a group of concepts that are used to represent vocabulary in the form of domain knowledge [18]. English lexicon enriches the vocabulary with equivalent words. Personality Detection is utilized to classify the features to different nine personalities. This phase applies a statistical approach. The proposed model uses Twitter as a social media platform. Twitter provides researchers with a rich source of text. Twitter presents a great opportunity to research human behavior in a natural context [19]. A case study is demonstrated to clarify that the results are similar to Enneagram expert analysis.

The proposed architecture contains multiple phases. The phases are text preprocessing, Word based feature extraction, word-based feature selection and personality detection. In text preprocessing, the text is cleaned from unnecessary elements like email, URL, numbers, stop words...etc. In feature extraction, the main function is to extract the word feature from text. In feature selection, the important features are elected according to the ontology and English lexicon. The last phase is personality detection. In this phase, the highest percentage of words which relate to a personality is chosen as the detective personality. Personality detection case study is applied on Twitter. A public person is analyzed by Enneagram experts. The proposed architecture also analyzes the same person's public Twitter 's profile. The case study uses the profile description and four tweets. The result of the case study illustrates that the output is similar to the Enneagram expert analysis.

This paper is organized into related work, personality detection architecture, personality detection case study, discussion and conclusion. Related work discusses past work of personality detection. Personality Detection Architecture describes the proposed one in detail. Personality Detection Case Study illustrates the applied case. Discussion section analyzes the results and its impact. Conclusion summarizes work done, infers results and demonstrates future work.

2. Related Work

Recent research work has applied different machine learning techniques. A proposed method extracted personality types from groups of essays using a convolutional neural network (CNN). The personality model was a big five personality model. They developed a novel document modeling technique based on a CNN features extractor [20]. This method includes data preprocessing, filtering, feature extraction and classification. In future they plan to incorporate more features, preprocessing and applying Long Short-Term Memory (LSTM) recurrent network to build both the sentence vector and the document vector They also plan to apply their document modeling technique to sentiment analysis or mood classification.

Most of the existing systems are mainly focused in one aspect of detecting the personality such as machine learning or using LIWC (Linguistic Inquiry and Word Count) features or using neural networks [21]. A proposed model that is composed of machine learning techniques, extracting LIWC features and Ontology-based personality detection. The personality model applied was the Big Five Model. Their future work is considering improving detection capacity in the consideration of texts and social behavioral aspects of a user on multiple social media.

Two types of techniques have been employed for detection of personality from the text i.e., machine learning-based approach and linguistic properties present in the text [3]. There are many used methods in personality traits from text. Some of these methods use a closed-vocabulary approach with psycholinguistic tools such as LIWC and others use an open-vocabulary approach by extracting n-grams and topics. Most personality prediction studies require a dataset to perform supervised learning, it is expensive to get the personality traits dataset of social media users. Further improvements to the existing state of personality prediction by applying more suitable algorithms or preprocessing methods to achieve higher accuracy, and implementing other personality models [22].

3. Personality Detection Architecture

The steps of personality detection are text prepossessing, word-based feature extraction, word-based feature selection and personality detection. Text prepossessing is used to cleanse the text from unnecessary extensions and non-useful text. Word-based feature extraction is used to take out features from the text as a starting phase. Word-based feature selection is applied to choose important features as a filtering phase. Personality Detection is utilized to classify the features to different nine personalities. The architecture is illustrated in figure 1.

3.1. Text Pre-processing

Data preprocessing is required to remove non-essential data such as URLs, Emails, symbols and others that are not important [23]. The target of the text pre-processing phase is to clean and prepare text for processing. The goal of the preprocessing procedure is to obtain significant information from the data [24]. The text pre-processing phase consists of hashtags removal, Twitter handler removal, URL removal, email removal, punctuation removal, special character removal, numbers removal, extra space removal, stop words removal, normalize Unicode, check spelling, stem words, lemmatize words and lower case folding as shown in figure 2.

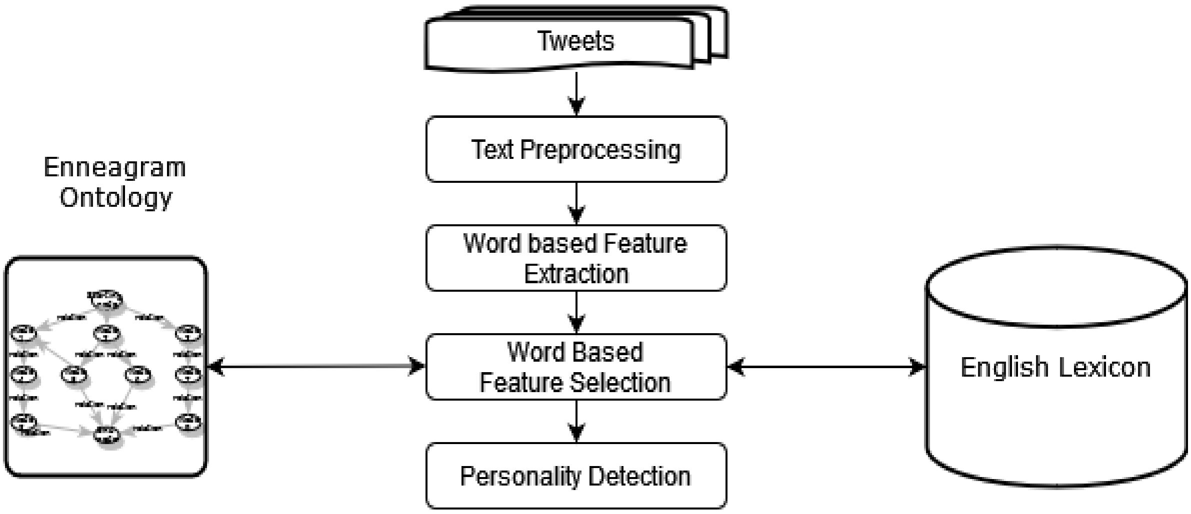


Figure. 1: Personality Detection Phases

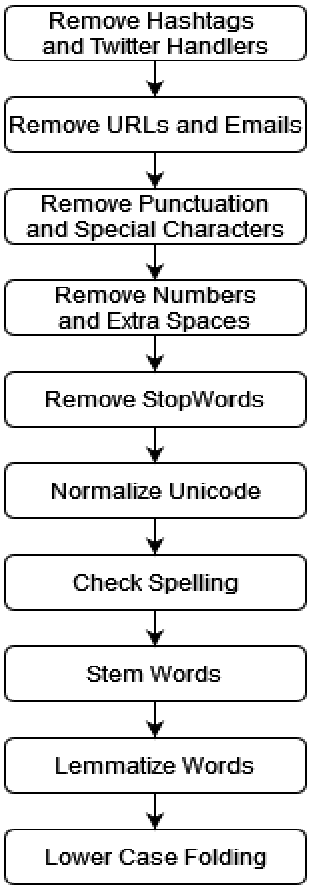


Figure. 2: Text Preprocessing Steps

3.2. Word-based Feature Extraction

The goal of this step is to extract the candidate features. The features include words. The preprocessed text is divided into tokens which is called tokenization. The tokens are converted to a bag of words as a feature representation. Word based feature extraction steps are demonstrated in figure 3.

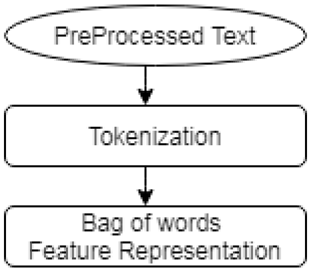


Figure. 3 word-based Feature Extraction

3.3. Word-based Feature Selection

Word-based Feature Selection uses both Enneagram ontology and English lexicon. The Enneagram ontology is utilized to represent nine personalities. The initial list of words is retrieved from the Enneagram Ontology. English Lexicon is required to find synonyms for the ontology's list of words. English Lexicon is also utilized to get antonyms of the fears. Words list is updated from the extracted terms of the English lexicon. Pre-processing Words are applied to the word list as a preparation phase. Features selection is based on a pre-processed updated list of words. Bag of words features are filtered according to this list as shown in figure 4.

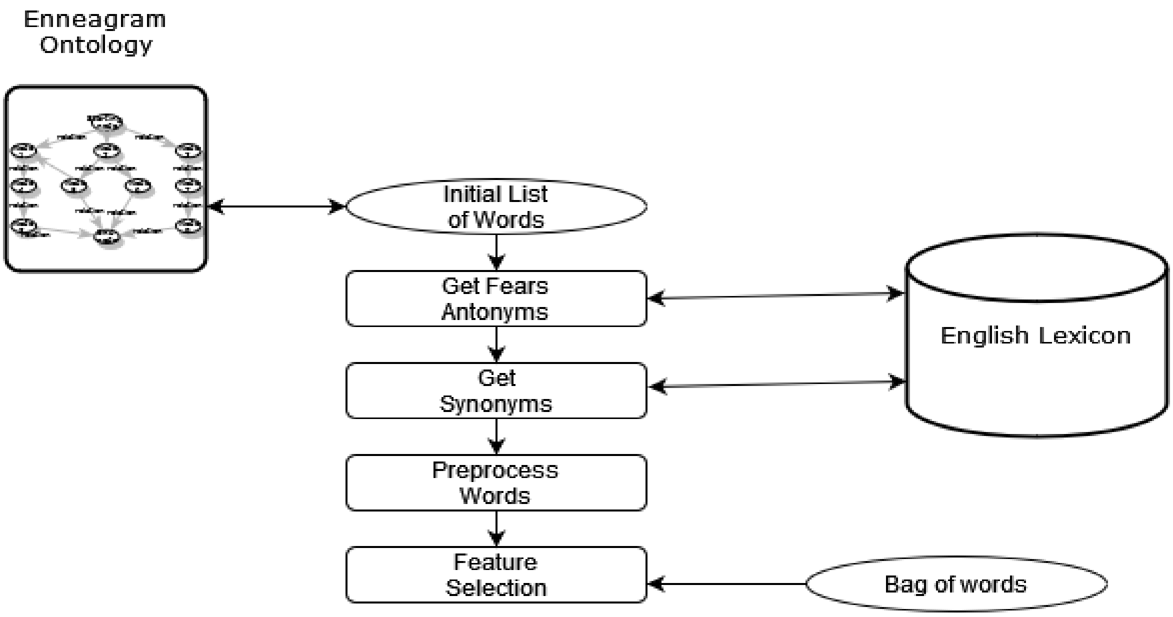


Figure. 4: word-based Feature Selection

3.4. Personality Detection

Personality Detection is utilized as a classification for the different nine personalities. After feature selection, there is a counting for each word occurrence in Twitter Text. These words are grouped together according to relevance to each personality. For each group of words, the total count of occurrence is calculated to know the distribution for each personality. Statistical technique is used to identify the personality type. Probability distribution is calculated by the occurrence number of features for each personality. The highest probability for a personality indicates that the personality is detected.

4. Personality Detection Case Study

Proposed architecture is applied to a small sample of Twitter text. This sample contains a description and four tweets from bill gates account on Twitter [25]. First the text preprocessing is used as a preparation phase. Before preprocessing, Twitter text is mentioned. After processing, Twitter text is cleaned using various methods like removing hashtags/handlers, URLS/Emails, punctuation/special characters, numbers/extra spaces, stop words....etc . After processing, The text is now ready for later steps. This phase is demonstrated as shown in table 1.

Table 1 Twitter Text Before and After Preprocessing

Type	Before PreProcessing	After Preprocessing
Description	"Sharing things I'm learning through my foundation work and other interests."	share thing im learn founda work interest
Tweet 1	"I was honored to speak at today’s \#LeadersClimateSummit about the three things we need to do to avoid a climate disaster. https://t.co/0YJJyabeIn "	honor speak today three thing need avoid climat dissat
Tweet 2	"Ambitious short-term goals like this are critical to moving closer to a net-zero future. As we rapidly scale the solutions we have, we must also invest in innovation to reach our ultimate goals. Thank you @POTUS for your leadership. https://t.co/SVeKp1KPHT"	ambiti shortterm goal like critic move closer netzero futur rapidly scale slut must also invest innov reach ultima goal thank leadership
Tweet 3	"It’s encouraging to see @POTUS Biden and @ClimateEnvoy Kerry re-establish America’s leading role on climate change. I look forward to joining leaders from around the world to talk about some of the most important challenges we need to overcome to avoid a climate disaster. https://t.co/a9CXZzIg6Y"	encourage see biden terri reestablish america lead role climat chang look forward join leader around world talk import challenge need overcome avoid climat dissat
Tweet 4	"Yesterday's verdict was a step in the right direction. But one court ruling alone will not bring to an end the injustice and inequity that Black people experience daily. I hope we will continue to make real progress on this."	yesterday verdict step right direct one court rule alon bring end injustice inequ black peopl experi daily hope continue make real progress

Second, the preprocessed text is tokenized into words. The words are chosen as features. The word features are represented as a bag of words. Bag of words model all words and the occurrence number for each word as shown in figure 5. The bag of words gives an insight like "climate" word is found in text three times; "thing" word is available in text two times...etc.

'climat': 3, 'thing': 2, 'need': 2, 'avoid': 2, 'dissat': 2, 'goal': 2, 'share': 1, 'im': 1, 'learn': 1, 'founda': 1, 'work': 1, 'interest': 1, 'honor': 1, 'speak': 1, 'today': 1, 'three': 1, 'ambiti': 1, 'shortterm': 1, 'like': 1, 'critic': 1, 'move': 1, 'closer': 1, 'netzero': 1, 'futur': 1, 'rapidly': 1, 'scale': 1, 'slut': 1, 'must': 1, 'also': 1, 'invest': 1, 'innov': 1, 'reach': 1, 'ultima': 1, 'thank': 1, 'leadership': 1, 'encourage': 1, 'see': 1, 'biden': 1, 'terri': 1, 'reestablish': 1, 'america': 1, 'lead': 1, 'role': 1, 'chang': 1, 'look': 1, 'forward': 1, 'join': 1, 'leader': 1, 'around': 1, 'world': 1, 'talk': 1, 'import': 1, 'challenge': 1, 'overcome': 1, 'yesterday': 1, 'verdict': 1, 'step': 1, 'right': 1, 'direct': 1, 'one': 1, 'court': 1, 'rule': 1, 'alon': 1, 'bring': 1, 'end': 1, 'injustice': 1, 'inequ': 1, 'black': 1, 'peopl': 1, 'experi': 1, 'daily': 1, 'hope': 1, 'continue': 1, 'make': 1, 'real': 1, 'progress': 1

Figure. 5: Case Study 1: Bag of Words

Third, the selected features use both Enneagram ontology and English lexicon. The occurrence number for each word is measured. Selected words-based features are also categorized according to personality type. Each personality type has related words which are found in a bag of words of the text. In investigator personality type, "innov" and "progress" words are found in text, the occurrence for each is one and one respectively. The highest personality percentage is the result which is 24 %. The result of this sample states that the personality is "Investigator" as shown in table 2. The result matches the analysis of the personality from Enneagram experts [26].

Table 2 Selected Features and Number of word occurrence for each personality type

Personality Type	Selected Features	Total
Investigator	’innov’: 1, ’progress’: 1, ’see’: 1, ’work’: 1, ’rule’:1, ’hope’: 1, encourage’:1, ’move’: 1, ’experi’: 1, ’share’: 1, ’continue’: 1, ’interest’: 1, ’like’: 1, ’step’: 1, ’chang’: 1, ’lead’: 1, ’invest’: 1, ’reach’: 1	24 %
Loyalist	’honor’: 1, ’challenge’: 1, ’encourage’: 1, ’lead’: 1, ’direct’: 1, ’avoid’: 2, ’step’:1, ’move’: 1, chang’: 1	13.33 %
Reformer	’critic’: 1, ’chang’: 1, ’right’: 1, ’black’: 1, ’real’:1, ’honor’: 1, ’lead’: 1, ’encourage’: 1	10.67 %
Individualist	’like’: 1, ’hope’: 1, ’thing’: 2, ’honor’: 1, ’direct’: 1, ’import’: 1, ’continue’: 1	10.67 %
Peacemaker	’hope’: 1, ’encourage’: 1, ’chang’: 1, ’continue’: 1, ’move’: 1, ’share’: 1, ’lead’:1, ’experi’: 1	10.67 %
Helper	’share’: 1, ’real’: 1, ’honor’: 1,, ’direct’: 1, ’import’: 1, ’need’: 2	9.33 %
Challenger	challenge’: 1, ’import’: 1, ’rule’: 1, ’direct’: 1, chang’: 1, ’move’: 1, ’continue’: 1	9.33 %
Achiever	’leadership’: 1, ’import’:1, ’ambiti’: 1, ’move’: 1, ’experi’: 1, ’honor’: 1	8 %
Enthusiast	’hope’: 1, ’work’: 1, ’thank’:1	4 %

5. Discussion

This case study states that using ontology, English lexicon and statistical approach can predict personality. The highest percentage of relevant keywords specify the person's directions, motivations, fears and desires. These calculations lead to inferring personality. The highest percentage is 24 %, which is the investigator personality. This concludes that the personality of the person is the investigator. The case study results match the Enneagram 's expert analysis. The result of this case study indicates that the author's text words can show the author's personality type. The author's text shows desires, motivations, characteristics and fears. Combination of ontology, English lexicon and statistical approach drives to good results. English lexicon enriches the ontology's keywords with synonyms. This drives the model to discover more relevant words related to personality. The statistical approach measures the probabilities for each personality. The highest probability of the output gives a signal to the direction of the author's personality.

This case study is a starting one which provides a positive direction. The current method to know Enneagram personality by using an assessment test. This proposed model is the first Enneagram personality detection system which gives a major advantage. Enneagram's personality detection system provides a fast identification. Unlike, assessment test which is long and takes time. Some patients may not be able to perform a test. The system aids psychiatrists and physicians to grant patients the right support in less time. Psychological support plays a great role in a patient's recovery. This assists patients to recover rapidly.

The limitation that this study is on a single account. These limitations cannot infer the precision of the proposed model. The case study is applied to a single personality. There is no clue about the effect of the model on other Enneagram's personalities. Outcomes required to do more results and case studies in order to infer the percentage of accuracy. The next step is to apply this architecture on many people which provides a more general view of the results. Investigating more accounts with different personalities is also a requirement.

6. Conclusion

Personality detection model is explained. The proposed model's four phases consist of text preprocessing, word-based feature extraction, word-based feature selection and personality detection. Text preprocessing removes unnecessary text. word-based Feature Extraction gets features from text. Word-based Feature Selection employs ontology and English lexicon to filter the features. Personality Detection applies a statistical approach to recognize personality.

The higher percentage indicates the person's personality. A case study of the proposed model is illustrated on Twitter text. The description and five tweets of a public profile are analyzed. The proposed model is applied on this sample. The highest percentage is the "Investigator" personality, which is 24 %. This outcome indicates that the profile personality is an investigator. The result is identical to the Enneagram experts' analysis from the official Enneagram institute. Enneagram is more in- depth in discovering personalities than other models. The traditional way to know the Enneagram is by using an assessment test. This model is the first automated personality detection that uses the Enneagram. This study has a positive consequence to understanding people deeply. This directs to help psychiatrists, physicians, relationships counseling and dating applications. The future direction is to apply the proposed model at a larger range and more profiles.

References

1. M. Kosinski, Y. Bachrach, P. Kohli, D. Stillwell, and T. Graepel, Manifestations of user personality in website choice and behavior on online social networks, Machine learning, vol. 95, no. 3, pp. 357–380, 2014.
2. O.-M. Sulea and D. Dichiu, Automatic profiling of twitter users based on their tweets, Working Notes Papers of the CLEF, 2015.
3. B. Agarwal, Personality detection from text: A review, International Journal of Computer System, vol. 1, no. 1, pp. 1–4,2014.
4. G. Farnadi, G. Sitaraman, S. Sushmita, F. Celli, M. Kosinski, D. Stillwell, S. Davalos, M.-F. Moens, and M. De Cock, Computational personality recognition in social media, User modeling and user-adapted interaction, vol. 26, no. 2, pp. 109–142, 2016.
5. W. Arthur, D. J. Woehr, and W. G. Graziano, Personality testing in employment settings: Problems and issues in the application of typical selection practices, Personnel Review, 2001.
6. A. Vinciarelli and G. Mohammadi, A survey of personality computing, IEEE Transactions on Affective Computing, vol. 5, no. 3, pp. 273–291, 2014.
7. D. Azucar, D. Marengo, and M. Settanni, Predicting the big 5 personality traits from digital footprints on social media: A meta-analysis, Personality and individual differences, vol. 124, pp. 150–159, 2018.
8. S. J. Guy, S. Kim, M. C. Lin, and D. Manocha, Simulating heterogeneous crowd behaviors using personality trait theory, in Proceedings of the 2011 ACM SIGGRAPH/Eurographics symposium on computer animation, 2011, pp. 43–52.
9. D. R. Riso and R. Hudson, The wisdom of the Enneagram: The complete guide to psychological and spiritual growth for the nine personality types. Bantam, 1999.
10. A. M. Bland, The enneagram: A review of the empirical and transformational literature, The Journal of Humanistic Counseling, Education and Development, vol. 49, no. 1, pp. 16–31, 2010.
11. A. Demir, O. Rakhmanov, K. Tastan, S. Dane, and Z. Akturk, Development and validation of the nile personality assessment tool based on enneagram, Journal of Research in Medical and Dental Science, vol. 8, no. 4, pp. 24–32, 2020.
12. R. Baron and E. Wagele, The Enneagram made easy: Discover the 9 types of people. Harper Collins, 2009.
13. M. Matise, The enneagram: An enhancement to family therapy, Contemporary Family Therapy, vol. 41, no. 1, pp. 68–78, 2019.
14. M. Alexander and B. Schnipke, The enneagram: A primer for psychiatry residents, American Journal of Psychiatry Residents' Journal, 2020.
15. V. Kaushal and M. Patwardhan, Emerging trends in personality identification using online social networks—a literature survey, ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 12, no. 2, p. 15, 2018.

16. E. Cambria, S. Poria, A. Gelbukh, and M. Thelwall, Sentiment analysis is a big suitcase, IEEE Intelligent Systems, vol. 32, no. 6, pp. 74–80, 2017.
17. N. F. Noy, D. L. McGuinness, et al., Ontology development 101: A guide to creating your first ontology, 2001.
18. W. Wu, H. Li, H. Wang, and K. Q. Zhu, Probase: A probabilistic taxonomy for text understanding, in Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, 2012, pp. 481–492.
19. A. Java, X. Song, T. Finin, and B. Tseng, Why we twitter: Understanding microblogging usage and communities, in Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, 2007, pp. 56–65.
20. N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, Deep learning-based document modeling for personality detection from text, IEEE Intelligent Systems, vol. 32, no. 2, pp. 74–79, 2017.
21. D. Sewwandi, K. Perera, S. Sandaruwan, O. Lakchani, A. Nugaliyadde, and S. Thelijjagoda, Linguistic features based personality recognition using social media data, in Technology and Management (NCTM), National Conference on IEEE, 2017, pp. 63–68.
22. V. Ong, A. D. Rahmanto, Williem, and D. Suhartono, Exploring personality prediction from text on social media: A literature review, Internetworking Indonesia, vol. 9, no. 1, pp.65–70, 2017.
23. M. D. N. Arusada, N. A. S. Putri, and A. Alamsyah, Training data optimization strategy for multiclass text classification, in 2017 5th International Conference on Information and Communication Technology (ICoIC7), IEEE, 2017, pp. 1–5.
24. A. Alamsyah, M. F. Rachman, C. S. Hudaya, R. P. Putra, A. I. Rifkyano, and F. Nurwianti, A progress on the personality measurement model using ontology based on social media text, in 2019 International Conference on Information Management and Technology (ICIMTech), IEEE, vol. 1, 2019, pp. 581–586.
25. Twitter bill gates, https://twitter.com/BillGates/ (visited on 29 May 2021).
26. Investigator official enneagram, https://www.enneagraminstitute.com/type-5 (visited on 29 May 2021)