

Received March 23, 2022, accepted April 27, 2022, date of publication May 5, 2022, date of current version May 19, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3172962

Assessing Information Quality of Wikipedia Articles Through Google's E-A-T Model

CHINTHANI SUGANDHIKA¹ AND SUPUNMALI AHANGAMA¹, (Member, IEEE)

Department of Information Technology, University of Moratuwa, Moratuwa 10400, Sri Lanka

Corresponding author: Supunmali Ahangama (supunmali@uom.lk)

This work was supported by the University of Moratuwa Senate Research Committee under Grant SRC/LT/2021/24.

ABSTRACT Along with the emergence of Web 2.0, User Generated Content (UGC) is becoming increasingly important for knowledge sharing. Wikipedia being the world's largest-ever community-based collaborative encyclopedia, is also one of the biggest UGC databases in the world. Wikipedia is dealing with a significant problem of Information Quality (IQ) because of its open-source and collaborative nature. When carrying out attacks such as link spamming, malicious users take advantage of Wikipedia's popularity on the World Wide Web (WWW). As a result, Wikipedia is generally not recommended for academic-related work. There are, however, some articles that are both rich in information and quality. Existing approaches for assessing Wikipedia's IQ involve statistical models and machine learning algorithms; however, the existing models do not produce satisfactory results. In this study, a novel theoretical model based on Google's E-A-T framework is introduced to assess Wikipedia's IQ. The model comprises three IQ constructs Expertise, Authority and Trustworthiness. Based on the empirical findings and study results, a set of IQ dimensions that influence the above three IQ constructs, as well as 45 IQ attributes to measure the IQ dimensions, were identified. The IQ attributes were automatically and inexpensively extracted from the content and meta-data statistics of Wikipedia articles using a Selenium 3.14 web automation script. A sample of 2000 articles comprising 1000 Featured Articles (FA) and 1000 non-FA articles from six WikiProjects was used for the data analysis. The proposed model was compared with three previously published models in terms of classification and clustering accuracy. It received classification and clustering accuracies of 95% and 93% respectively, which is a drastic improvement over the existing models. Furthermore, an average inter-rater agreement of 84% was observed. Thus, the proposed model's effectiveness is fairly validated by this extensive experiment. This study contributes to the related knowledge area by introducing a novel framework to assess Wikipedia articles' IQ.

INDEX TERMS Authority, expertise, Google's E-A-T, hybrid approach, information quality, trustworthiness, web 2.0, Wikipedia.

I. INTRODUCTION

Wikipedia stands as the world's largest and the most popular community-based collaborative multilingual encyclopedia to date. As of October 2021, it had over 54 million articles in over 300 languages [1]. It is a well-known fact that Wikipedia has become the default site for many internet users looking for information. The relevant domain experts create the content of traditional encyclopedias, and therefore, the Information Quality (IQ) of the contents is to the highest standards. The content of collaborative platforms like Wikipedia, on the

other hand, is being improved with the help of numerous contributors. This group of contributors can be either registered users or anonymous users. As a result, Wikipedia's IQ has always become an open topic of discussion.

Malicious users use the popularity of Wikipedia on the World Wide Web (WWW) and its open-model concept as advantages when mounting attacks [2]. Wikipedia vandalism is one such issue. The Wikipedia community has defined vandalism as "any change made to content to compromise its integrity" [3]. Insertion of obscenities and crude humor within the content is one such major attack. This inversely affects not only Wikipedia's integrity but the image of the relevant personnel or the organization which the content is

The associate editor coordinating the review of this manuscript and approving it for publication was Fu-Kwun Wang¹.

about [4]–[6]. Another type of attack is link spamming, which is defined as the placement of inappropriate and promotional external links in Wikipedia materials [2]. It can lead significant traffic to a landing site and result in monetary benefits to the spammer. Other sorts of vandalism include blanking the page or removing the content, changing the article's formatting, and adding offensive edit summaries. Furthermore, certain facts are not cited properly, the references' destination links are broken or unavailable and, images, graphs, and maps placed within the content are not copyrighted.

As a result, academics advise students not to use Wikipedia as a research or academic information source [7], [8]. Furthermore, several universities and secondary schools have policies prohibiting students from using Wikipedia in their academic writings [9]. It is worth noting, however, that not all of the Wikipedia contents are unreliable or lacking in quality standards. Even if Wikipedia suffers from IQ issues in the big picture, certain articles are rich in information and quality. Due to this overall IQ issue of Wikipedia, such quality and credible articles are also missed out.

A. WIKIPEDIA'S INFORMATION QUALITY

The term information is defined as “useful data that have been processed in such a way as to increase the knowledge of the person who uses the data” [10]. As defined by Gustavsson and Wänström [11], IQ is the “ability to satisfy stated and implied needs of the information consumer”. In other words, it is the “measure of the value which the information provides to the user of that information” [12]. IQ is perceived subjectively. Britannica [13] has defined the encyclopedia as “a reference work that contains information on all branches of knowledge or that comprehensively treats a particular branch of knowledge”. Wikipedia is a collaborative community-based encyclopedia. Unlike a typical encyclopedia, which is compiled by domain experts, a community-based encyclopedia is compiled by a broad group of people who may or may not be domain experts. It is, therefore, challenging to produce accurate and reliable content in a collaborative encyclopedia. Accordingly, the IQ of community-based collaborative encyclopedias could be defined as providing more reliable, accurate and comprehensive information on many knowledge areas in a way that a layman can comprehend and understand.

The English Wikipedia currently has 6,291,946 articles, with an average of 594 new articles every day [14]. On the other hand, human judgment based on a limited number of experts is used to determine quality on Wikipedia. This existing quality grading system tags the articles into seven ordinal classes as Featured Articles (FA), A-Class, Good Articles (GA), B-Class, C-Class, Start, and Stub classes from best to worst [15]. FA class is considered the highest quality class among these classes. However, it is obvious that the community cannot constantly monitor and evaluate the quality of Wikipedia articles against numerous edits and revisions made every day. As a result, a fully/semi-automated

method of evaluating IQ in Wikipedia is required to facilitate Wikipedia's collaboration.

Prior studies in [16], [17] have presented diverse quality determinant factors and theoretical frameworks for assessing Wikipedia content's IQ. However, the complexity and preciseness of certain frameworks and algorithms do not deliver sufficient classification performance. Certain methods are not automated. Especially, the data collection phase requires a substantial amount of time and effort. Furthermore, proper attempts to investigate how certain facts in Wikipedia articles affect IQ, such as content authority, verifiability, and maturity, are lacking [18]. Therefore, this research aims to introduce a novel theoretical model based on Google's E-A-T quality framework [19] to better classify Wikipedia articles by quality. Based on an extensive literature review and evaluation process, authors define E-A-T: i.e., Expertise, Authority and Trustworthiness as the key constructs in assessing the IQ of Wikipedia articles. Thus, this study attempts to understand (1) influencing IQ factors for the E-A-T model and (2) the reasons for getting a better quality assessment for Wikipedia using the E-A-T model compared to previous algorithms.

Accordingly, this paper's major contributions can be summarized as follows. Firstly, the study introduces a novel theoretical model based on Google's E-A-T framework to assess the IQ of Wikipedia. Expertise, Authority and Trustworthiness are presented as the novel constructs for Wikipedia's IQ assessment. Secondly, a set of IQ dimensions that influence the E-A-T constructs was identified. Thirdly, a comprehensive set of 45 IQ attributes (14 new and 31 existing attributes) were presented to measure each dimension. To extract features from Wikipedia, an efficient automated web scraping strategy was used. The scraping strategy is hybrid, where the features were extracted from both article content and meta-data. The dataset was analyzed using a statistical analysis followed by a machine learning approach. A total of 2000 articles were extracted from six WikiProjects for the dataset. Afterwards, the model was thoroughly evaluated in various ways to ensure its validity. Assessing the IQ of Wikipedia articles would be valuable for both readers and collaborators (reviewers, authors and editors). Since Wikipedia publishes millions of articles, it might be difficult for readers to extract high-quality information. Therefore, directing people toward high-quality articles would be beneficial. Furthermore, because manually assessing each article is impractical, reviewers would be benefited from an automated process of identifying flows in the articles. As a result, this would aid in achieving more efficient knowledge collaboration and influencing the role of online communities.

The rest of the paper is organized as follows. Section 2 is about the literature review. The adopted methodology is explained in Section 3. Section 4 is about data analysis and results, and Section 5 is about Evaluation. Finally, Section 6 and Section 7 are about the discussion and conclusion of the study, respectively.

II. LITERATURE REVIEW

The literature review is discussed under existing approaches for IQ assessment of Wikipedia and the chosen theoretical model, i.e., E-A-T.

A. EXISTING APPROACHES FOR WIKIPEDIA IQ ASSESSMENT

Literature has presented three major approaches for determining Wikipedia's IQ. These are (1) Meta-data-based approach - exploits the meta-information of the articles such as edits, editors and revision history. Studies presented in [20]–[24] are some examples; (2) Content-based approach - which considers the content related features such as article length in words [25]–[28]; (3) Hybrid approach – a combination of the above two approaches. Studies such as [17], [29]–[31] have adopted this hybrid approach. Based on these approaches, various theoretical frameworks, formulas, and statistical models have been proposed to measure Wikipedia's IQ. Table 1 describes these previous work and their limitations. Most of these approaches require extensive human effort in the feature extraction process. However, machine learning models, such as K-Nearest Neighbour (KNN) and Support Vector Regression (SVR) have been adopted by [29], [32], [33].

IQ is a multi-dimensional construct [34], [35] and thus, aspects of IQ can be split into various dimensions which can be measurable [13]. Accordingly, a set of measurable quality attributes together represent a single IQ dimension. Drawing from Google's E-A-T model and following the hybrid approach of article content and meta-data, the authors reviewed the accumulated literature to understand various dimensions and attributes adopted by previous studies in assessing Wikipedia's IQ.

B. THEORETICAL FRAMEWORK - E-A-T MODEL

Since Google's algorithm update in August 2018, the E-A-T quality framework has been a major focus of attention. The model consists of three constructs E-Expertise, A-Authority and T-Trustworthiness. Expertise is defined as the knowledge and skills of the Main Content (MC) creator. Authority is the "authoritativeness of the creator of the MC, the MC itself, and the website", and Trustworthiness is defined as "trustworthiness of the creator of the MC, the MC itself, and the website" [19]. Thus, through the E-A-T framework, Google aims to deliver the highest quality content and experience to the online search community. Since Wikipedia is an online source of information, authors believe that E-A-T is a more appealing model for assessing the IQ of collaborative platforms such as Wikipedia. None of the previous studies has adopted E-A-T as the base model to conduct a systematic review of online collaborative content IQ assessment to the best of the authors' knowledge.

1) EXPERTISE

A high-quality website requires enough expertise to be authoritative and trustworthy on the respective topics.

An "Expert Wikipedia article" is exclusively knowledgeable and informative and is composed of contributors with relevant expertise. Prior studies have attempted to evaluate IQ through expertise. For example, certain studies in [20]–[22], [36], [37] have analyzed how various editor characteristics such as editor experience, concentration, coordination and coordination patterns of editors impact the quality of articles on Wikipedia. Studies such as [38], [39] have adopted graph techniques and networking concepts to identify the behavior of editors. The studies in [25], [28] argue that article length is a precise measure of article quality despite its simplicity. Some studies have focused on article lifecycle-based approaches [40]. Further, Lipka and Stein [27] have analyzed the articles' writing styles. The frameworks introduced in [16], [17], [41] included a separate indicator for informativeness. The term "informativeness" can be defined as the ability to provide useful information and comprehensive content. Graphical representation of information improves informativeness [42], [43]. Authors in [44] have analyzed the language quality of Wikipedia. Studies in [45]–[47] have adopted readability metrics. Readability scores reflect the easiness to read a given text. From a general encyclopedia's perspective, the information presented should be easily understood by a layperson [13]. Other than Wikipedia related studies, studies such as [48] also have used readability as an index for assessing the quality of web content.

2) AUTHORITY

When a source of knowledge is cited by others more frequently, it becomes more reputable and authoritative within its verticals [49]. When a source is more authoritative, it implies it has a higher level of quality. The authoritativeness of the contributors, the content, and the website itself may all contribute to the authority of a Wikipedia page. Certain studies have attempted to assess IQ based on the authors' reputation. Studies in [5], [23] have found that high-rated articles have been authored and edited by highly reputed authors and editors. Similarly, Nemoto *et al.* [24] have found that articles initiated by reputed authors rapidly progress to a high-quality status. However, there is a lack of studies that have focused on the authority of content and websites in evaluating Wikipedia's IQ. Only Lewandowski and Spree [56] have adopted Wikipedia article rankings in search engines to assess IQ.

3) TRUSTWORTHINESS

The term "reliability" refers to the ability to provide accurate and trustful information. In collaborative contexts, reliability and trustworthiness are important factors. A trustworthy Wikipedia article contains factually accurate and verifiable information. Zeng *et al.* [57] have proposed a revision history-based technique for evaluating an article's trustworthiness. They have created a Bayesian network trust model based on article revision history information. Halfaker and Taraborelli [6] have created an online Objective Revision Evaluation Service (ORES). The single input parameter for

TABLE 1. Previous Studies about IQ Assessment of Wikipedia.

Approach	Previous work	Main focus / limitations
Meta-data based approach – Exploits meta-data of the articles	Harnessing the wisdom of crowds in Wikipedia: quality through coordination – 2007 [20], Does it matter who contributes? – A study on featured articles in the German Wikipedia – 2008 [23], Wikipedia vandalism detection: combining natural language, meta-data, and reputation features – 2008 [5], Quality assessment of Wikipedia articles using h-index – 2015 [50], Automatically assessing Wikipedia article quality by exploiting article–editor networks – 2015 [51]	These studies in [5], [20], [23] have mainly focused on edit history, concentration and coordination of the contributors in assessing the IQ of Wikipedia articles. Authors in [50] have used editor quality as a quality indicator. The study in [51] has used an article–editor network and PageRank based models. However, this approach requires collecting large amounts of meta-data about the articles. Further, the IQ prediction is indirect since it is based on external factors such as contributors rather than considering the article content-related factors [42].
Content-based approach – Exploits article internal features	Size matters: Word count as a measure of quality on Wikipedia – 2008 [25], Measuring article quality in Wikipedia: Lexical clue model – 2011 [52], Predicting quality flaws in user-generated content: The case of Wikipedia – 2012 [18], Predicting quality flaws in user-generated content: The case of Wikipedia – 2011 [18], Tell me more: An actionable quality model for Wikipedia – 2013 [16], The success and failure of Quality Improvement Projects in peer production communities – 2015 [41], An end-to-end learning solution for assessing the quality of Wikipedia articles – 2017 [32], Content matters: clustering web pages for QoE analysis with WebCLUST – 2021 [53]	This approach assesses IQ using the features derived from the article content itself. Article length measured using word count is a simple yet precise measure for IQ assessment of Wikipedia [25]. A high accuracy has been achieved in classifying the articles into FA and non-FA articles using various lexical features [52]. The study in [27] has used the article's writing style by exploiting the binarized character trigram features as an IQ determinant factor. Furthermore, features such as completeness, informativeness, number of headings, number of images, references and readability scores also have been used [16], [18], [32], [41]. Other than Wikipedia, certain studies have found that web page content positively impacts the web user's quality of experience [53].
Hybrid approach – Exploits meta-data and article content related features	Information quality work organization in Wikipedia – 2008 [17], Automatic quality assessment of content created collaboratively by web communities: A case study of Wikipedia – 2009 [29], Relative quality and popularity evaluation of multilingual Wikipedia articles – 2017 [43], Multilingual ranking of Wikipedia articles with quality and popularity assessment in different topics – 2019 [54], Quality and importance of Wikipedia articles in different languages – 2016 [55], Artificial Intelligence service 'ORES' gives Wikipedians X-ray specs to see through bad edits – 2017 [6], Assessing the quality of information on Wikipedia: A deep-learning approach – 2019 [30], NwQM: A neural quality assessment framework for Wikipedia – 2020 [31]	This approach adopts both meta-data and article internal features to assess the IQ of Wikipedia. The pioneering study related to this approach was presented in [17]. Informativeness, Completeness, Complexity, Consistency, Currency, Volatility, Revision History, Popularity, Citation Network and Readability are some of the Hybrid IQ measures introduced under this approach [31], [48], [29]–[55]. Moreover, Objective Revision Evaluation Service (ORES) is a service that automates the tasks of vandalism detection and removal of untrustful edits. Additionally, for certain language versions ORES evaluates articles on a scale between 0 and 1 (currently supports only 9 languages) [6].

ORES is the revision ID of a Wikipedia article. The authors in [18] argued that the article's content should always be verifiable, which indicates that a reader should be able to confirm the accuracy of information regardless of the authority of the source. Nicholson *et al.* [58] tried to evaluate Wikipedia articles' reliability through the references. Lewoniewski *et al.* [43] have conducted an interesting study regarding the popularity and reliability modelling. Moreover, in the hybrid frameworks introduced in [16], [17], [41] separate indicators have been adopted to indicate the reliability of the articles. Besides that, these hybrid studies have presented comprehensive features including content, network, structure and edit history, style, readability and review features, that can be extracted from both content and meta-data statistics [18], [59]. However, most of those features are hand-engineered features.

According to the above review, the notable limitations in this area of knowledge are as follows. First, the authority has not been considered a prominent indicator in assessing the IQ of Wikipedia. Higher web presence leads to a higher web reputation due to other experts or influencers citing the specific website/page as a source of information. Correspondingly, a good reputation improves the content's quality [60], [43]. Furthermore, the authors observed a drastic reduction in the classification accuracy (accuracy reduced from 95% to 79%) when the authority component was removed from the proposed model. Therefore, the authority of the presented content should be considered a key factor in assessing Wikipedia's IQ. Second, based on the literature, there is a dearth of studies on how verifiability and maturity of the content affect the IQ [18]. Finally, the previously presented models and algorithms in [17], [21], [42] do not provide satisfactory classification performance in classifying the articles into correct quality classes [61]. Considering all these facts, it can be concluded that the Expertise, Authority and Trustworthiness of Wikipedia articles precisely represent Wikipedia's IQ. Therefore, it is notable that Google's perception through the E-A-T model can be drawn to assess the IQ. Aligning with the empirical findings, the authors present Informativeness, Readability and Understandability as the dimensions that influence Expertise and Maturity, Verifiability and Reliability as the dimensions that influence Trustworthiness. Moreover, the authors present 45 IQ attributes that can be automatically and inexpensively extracted from the content and meta-data of the articles (see Appendix A). The conceptual framework, operationalization of the above IQ dimensions (Informativeness, Readability Understandability, Reliability, Verifiability and Maturity) using attributes and thus, synthesizing the three constructs: Expertise, Authority and Trustworthiness, is presented in the next sections.

III. METHODOLOGY

A. DATA COLLECTION

This study is quantitative research where statistical analysis and experimentation were used to analyse data. The

IQ attributes were extracted from the articles' content and meta-data about the articles. Therefore, the data collection technique was secondary. An algorithm was written using Selenium 3.14 to automatically and inexpensively scrape these features from articles using their Uniform Resource Locators (URL). Selenium is a cross-browser web application testing and validation framework which is free and open-source [62]. The majority of previous studies were based on manual feature extraction approaches, which are time-consuming [42]. To overcome these limitations, this research proposes an efficient and inexpensive data gathering technique. This technique can be used to assess the quality of not just Wikipedia but also other online collaborative repositories.

The articles in the Wikipedia repository form the study's population. Wikipedia has categorized articles into WikiProjects that share a common knowledge base, such as Medicine and Politics, History, Geography, Sports, Plants and Computer Science, etc. Accordingly, for the data analysis, articles from six WikiProjects Medicine [63] Politics [64], Sports [65], History [66], Science [67] and, Biographies [68] were chosen. These WikiProjects were chosen because they are considered the most saturated and active WikiProjects [63]. They contain a significant number of FA articles that have undergone an extensive review process. Even if almost all the WikiProjects contain plenty of non-FA articles, most of the WikiProjects contain fewer FA articles. For example, WikiProject Medicine contains only 62 FA articles. Authors had to collect 1000 FA articles and these six WikiProjects helped us to collect this number of FA articles. Aligning with the community assumption about the FA articles' high quality, the authors utilized FA as the basis for assessing the IQ of Wikipedia articles [16], [17]. The sample comprised 2000 articles, including 1000 FA and 1000 non-FA articles extracted from the above six WikiProjects. Then, the 45 IQ attributes (see Appendix A) were extracted from each of these 2000 articles. Thus, the unit of the analysis was a single article.

B. CONCEPTUAL FRAMEWORK AND OPERATIONALISATION OF THE CONSTRUCTS

Based on the E-A-T framework and empirical findings on Wikipedia IQ assessment, the study proposes the conceptual framework given in Figure 1. According to the conceptual framework, the study defines three IQ constructs, a set of IQ dimensions that influence the constructs and 45 attributes to measure each dimension. The three constructs are (1) Expertise, (2) Authority and (3) Trustworthiness. The dimensions of Expertise are (1) Informativeness, (2) Readability and (3) Understandability. The dimensions of Trustworthiness are (1) Reliability (2) Verifiability and (3) Maturity.

Following the data collection phase, a statistical analysis was conducted using SPSS 26.0. SPSS was specifically

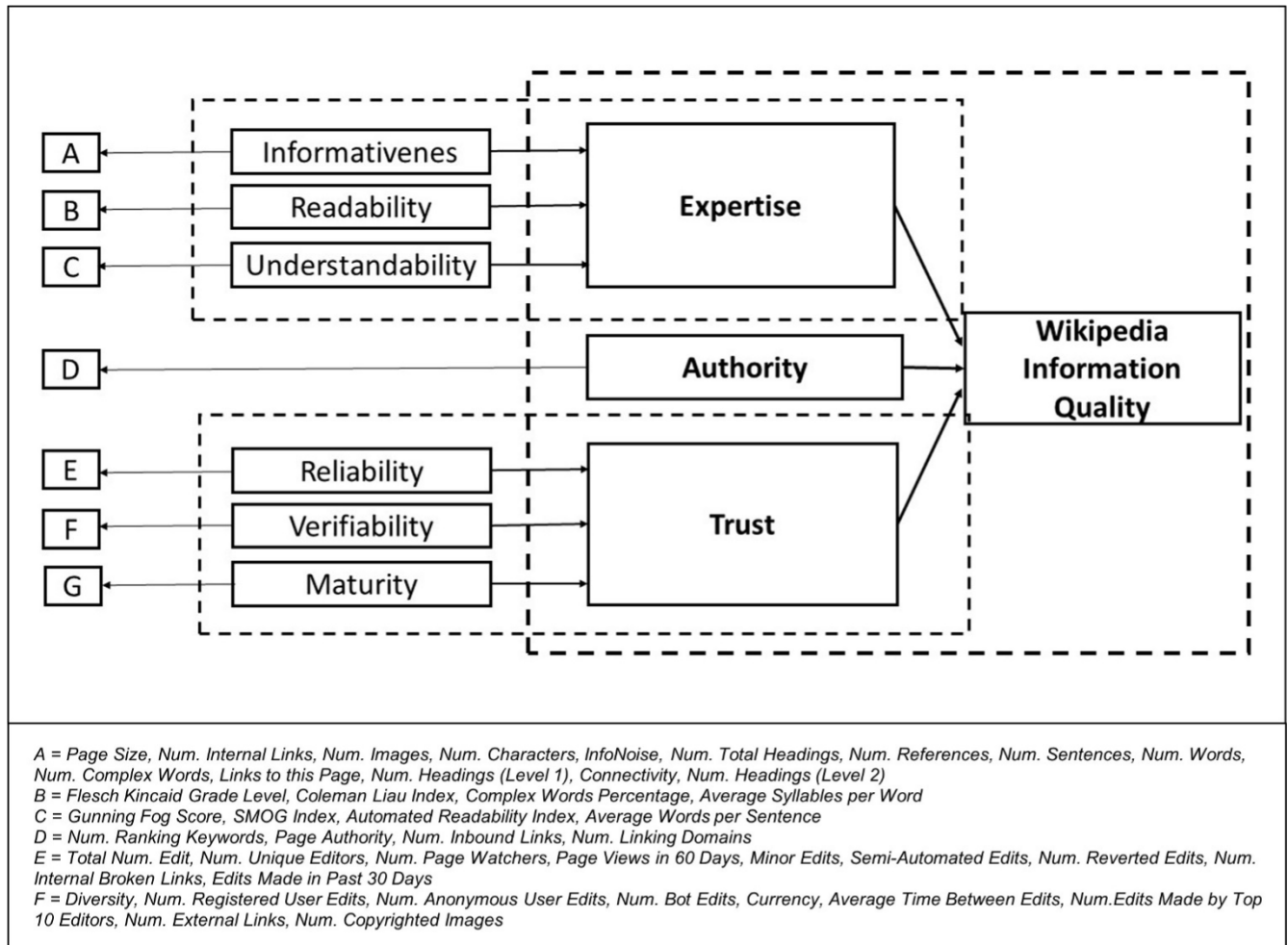


FIGURE 1. Proposed conceptual framework and operationalization of the study variables.

used because, of its easiness to use and ability to process critical data in simple steps. For the evaluation, a machine learning-based approach was adopted. The algorithms Decision Tree, Logistic Regression, Support Vector Machine (SVM), Naïve

Bayes, KNN and K-Means were used because these algorithms are capable of handling both numerical and categorical data. Further, these algorithms have been adopted by previous studies to assess the IQ of Wikipedia [17]. Initially, an Exploratory Factor Analysis (EFA) analysis was conducted to discover the factor structure of Wikipedia's IQ and to examine its internal reliability. Traditionally, EFA has been used to explore the possible underlying latent structure of a set of observed variables without imposing a preconceived structure on the outcome whilst Confirmatory Factor Analysis (CFA) has been used to confirm the latent structure [69]. Since it was required to identify the possible factor structure for the 45 IQ attributes, an EFA was conducted. This methodology has been adopted by previous well-known studies related to this knowledge area such as [17] and therefore, authors followed this procedure when defining the IQ

constructs and IQ dimensions. By conducting two EFAs, the IQ dimensions that influence each of the constructs were identified, and the seven dimensions were operationalized using the 45 IQ attributes. In Figure 1, A to G indicate the relevant IQ attributes that affect each IQ dimension. Accordingly, seven metrics (1st order IQ functions) to measure the IQ dimensions and three metrics (2nd order IQ functions) to measure the E-A-T constructs were derived.

Subsequently, the authors conducted a regression analysis to observe the association of each quality construct with IQ. The suggested model's performance was then evaluated using three approaches. They are (1) classification performance, (2) clustering performance and (3) Fleiss-Davies Kappa interrater reliability test [70].

IV. DATA ANALYSIS AND RESULTS

The dataset was then subjected to a cleaning process using IBM SPSS software. When extracting meta-data from the articles, several values (0.8%) were empty for some IQ attributes, and therefore, a series mean was used to replace the missing values to avoid loss of data points [71].

A. 1ST ORDER IQ FUNCTIONS

The first EFA was conducted for the 45 IQ attributes to identify the underlying structure as the first step. The analysis suggested seven components. Accordingly, the following seven 1st order IQ functions were defined to derive the values for each of the IQ dimensions based on the nature of the suggested variable groupings and empirical findings (refer to equations (1)-(7)). Since the IQ measurements were not standardized and employed various scales at the beginning, the authors retained the component score coefficients when defining the functions [9].

$$\begin{aligned} \text{Informativeness} = & 0.12 * \text{Page Size} + 0.164 \\ & * \text{Num. Internal Links} + 0.12 \\ & * \text{Num. Images} + 0.182 \\ & * \text{Num. Characters} + 0.181 \\ & * \text{InfoNoise} + 0.17 \\ & * \text{Num. Total Headings} + 0.141 \\ & * \text{Num. References} + 0.151 \\ & * \text{Num. Sentences} + 0.154 \\ & * \text{Num. Words} + 0.155 \\ & * \text{Num. Complex Words} + 0.016 \\ & * \text{Links to this Page} + 0.016 \\ & * \text{Connectivity} + 0.11 \\ & * \text{Num. Headings (Level 1)} + 0.11 \\ & * \text{Num. Headings (Level 2)} \quad (1) \end{aligned}$$

$$\begin{aligned} \text{Readability} = & 0.213 * \text{Flesch Kincaid Grade Level} \\ & + 0.185 * \text{Coleman Liau Index} + 0.26 \\ & * \text{Complex Words Percentage} + 0.253 \\ & * \text{Average Syllables per Word} \quad (2) \end{aligned}$$

$$\begin{aligned} \text{Understandability} = & 0.393 * \text{Gunning Fog Score} + 0.352 \\ & * \text{SMOG Index} + 0.181 \\ & * \text{Automated Readability Index} + 0.344 \\ & * \text{Average Words per Sentence} \quad (3) \end{aligned}$$

$$\begin{aligned} \text{Authority} = & 0.253 * \text{Page Authority} + 0.349 \\ & * \text{Num. Linking Domains} + 0.307 \\ & * \text{Num. Inbound Links} + 0.349 \\ & * \text{Num. Ranking Keywords} \quad (4) \end{aligned}$$

$$\begin{aligned} \text{Reliability} = & 0.114 * \text{Total Num. Edits} + 0.1 \\ & * \text{Num. Unique Editors} + 0.176 \\ & * \text{Num. Page Watchers} + 0.167 \\ & * \text{Page Views in 60 Days} + 0.107 \\ & * \text{Minor Edits} + 0.119 \\ & * \text{Semi - Automated Edits} + 0.129 \\ & * \text{Num. Reverted Edits} + 0.119 \\ & * \text{Num. Internal Broken Links} + 0.119 \\ & * \text{Edits Made In Past 30 Days} \quad (5) \end{aligned}$$

$$\begin{aligned} \text{Verifiability} = & 0.526 * \text{Diversity} - 0.184 \\ & * \text{Num. Registered User Edits} + 0.144 \\ & * \text{Num. Anonymous User Edits} + 0.135 \\ & * \text{Num. Bot Edits} + 0.464 * \text{Currency} \\ & + 0.36 * \text{Average Time Between Edits} \\ & - 0.085 \\ & * \text{Num. Edits Made by Top 10 Editors} \\ & - 0.144 * \text{Num. External Links} + 0.45 \\ & * \text{Num. Copyrighted Images} \quad (6) \end{aligned}$$

$$\begin{aligned} \text{Maturity} = & 0.561 * \text{Median Revert Time} \\ & + 0.522 * \text{Age} \quad (7) \end{aligned}$$

B. 2ND ORDER IQ FUNCTIONS

After identifying the variable grouping through the first EFA, the dataset was tested for the four parametric assumptions since these metrics are subjected to regression analysis. These assumptions are normality (Skewness and Kurtosis were above and below zero), linearity (all the scatter plots showed a linear behavior), multicollinearity (Variance Inflation Factor (VIF) was less than 5), and homoscedasticity (residuals plots with regression line method were used and significant homoscedasticity was observed in each measured variable). The descriptive statistics of 1st and 2nd order IQ functions are given in Table 2. Satisfied four assumptions revealed that data are ready for further analysis [72].

Next, the dataset with values for all the IQ dimensions was fed to the second EFA to identify the IQ dimensions affecting each E-A-T construct. The analysis exactly suggested three components as the authors expected. The three constructs, Expertise, Authority and Trustworthiness, were derived by observing the nature of the factor loading along with the findings from the literature. Accordingly, the IQ dimensions that affect the E-A-T model were identified (see Figure 1). The following 2nd order IQ functions were defined to measure each construct (refer to equations (8)-(10)). For both EFAs, Principal Component Analysis (PCA) was used as the extraction method, and Varimax with Kaiser Normalization was used as the rotation method.

$$\begin{aligned} \text{Expertise} = & 0.255 * \text{Informativeness} + 0.654 \\ & * \text{Readability} + 0.557 * \text{Understandability} \quad (8) \end{aligned}$$

$$\text{Authority} = 0.37 * \text{Authority} \quad (9)$$

$$\begin{aligned} \text{Trustworthiness} = & 0.362 * \text{Reliability} + 0.369 * \text{Maturity} \\ & - 0.846 * \text{Verifiability} \quad (10) \end{aligned}$$

C. MULTIPLE LINEAR REGRESSION MODEL

To study the nature of the association between each of the quality constructs (i.e., Expertise, Authority and Trustworthiness) with IQ, a multiple linear regression analysis was conducted. The dependent variable is IQ. It is categorical and

TABLE 2. Descriptive Statistics of the IQ Dimensions and IQ Constructs.

IQ dimension	Mean	SD	Skewness	Kurtosis	VIF	IQ constructs	Mean	SD	Skewness	Kurtosis	VIF
Informativeness	3.87	0.91	0.05	-1.37	1.32	Expertise	2.08	0.28	-0.21	-1.03	1.67
Readability	0.77	0.07	0.11	2.83	1.53						
Understandability	0.98	0.14	-0.97	2.84	1.89						
Authority	2.36	0.84	0.53	-0.26	1.57	Authority	2.97	0.31	-0.53	-0.26	1.39
Reliability	1.76	0.73	0.39	-1.07	1.25	Trustworthiness	3.23	0.46	0.10	0.26	1.32
Verifiability	1.65	0.46	-0.56	1.37	1.81						
Maturity	1.06	0.90	-0.33	0.06	2.18						

SD – Standard Deviation; VIF - Variance Inflation Factor

reflects the relevant quality grading class for the article. The two classes are the FA class and the non-FA class. The three IQ constructs are the independent variables. The numerical values for the three independent variables were derived by substituting the values of IQ attributes using the above 1st order and 2nd order IQ functions. Accordingly, the following multiple linear regression model (refer to equation (11)) was derived.

$$\begin{aligned}
 \text{Information Quality} = & 1.191 * \text{Expertise} + 0.281 * \text{Authority} \\
 & + 0.165 * \text{Trustworthiness} - 2.139
 \end{aligned}
 \tag{11}$$

The R² value of 0.75 indicates that the regression model has high explanatory power. Thus, it can be concluded that Expertise, Authority and Trustworthiness reasonably predict IQ. According to the results, it was observed that the Expertise of a Wikipedia article is positively associated with its IQ (β = 0.673, p< 0.000). Similarly, the Authority of a Wikipedia article is positively associated with its IQ (β = 0.175, p< 0.000), and the Trustworthiness of a Wikipedia article is also positively associated with its IQ (β = 0.154, p< 0.000) (refer to Table 3). Thus, the study results also support that Expertise, Authority and Trustworthiness improve Wikipedia articles’ IQ.

V. EVALUATING THE PROPOSED MODEL

The proposed model was evaluated and validated in three ways, (1) Classification performance, (2) Clustering performance and (3) Interrater-reliability test results.

A. CLASSIFICATION PERFORMANCE

The model’s classification performance was tested using the same dataset used in the study as the first evaluation method.

The independent variables were Expertise, Authority and Trustworthiness, while the dependent variable was IQ.

TABLE 3. Results of the Regression Analysis.

Independent variable	β
Expertise	0.673***
Authority	0.175***
Trustworthiness	0.154***

N = 2000; *p<0.05; **p<0.01; ***p<0.001 (2-tailed); R = 0.886; R²=0.750; Adjusted R²= 0.747

Stvilla *et al.* [17] have tested their model using the Decision Tree classifier. In the Decision Tree, the tree could be inspected to learn how specific features influence building complex models. Therefore, the authors also adopted the same classifier to assess the proposed model’s classification performance in terms of accuracy and F1 score. The accuracy is the fraction of correct predictions, and the F1 score combines the precision and recall into a single metric by considering that there is a harmonic mean. Accordingly, the proposed model received a 95% accuracy and a 94% of F1 score.

Consequently, the suggested model’s classification performance was compared to that of the previous three well-known models. Stvilla *et al.* [17] proposed Model 1 in 2005 and is regarded as the pioneering research work related to this knowledge area. Model 2 was presented in 2013 by Warncke-Wang *et al.* [16] and Model 3 is a deep learning-based model introduced in 2017 by Shen *et al.* [42]. Shen *et al.* [42] formulated the quality assessment as a classification problem and proposed a Long Short Term Memory (LSTM) based model using a set of hand-engineered features. This model has achieved 6.5% higher accuracy than

TABLE 4. Comparison of the Proposed Model's Classification Performance with Model1, Model 2 and Model 3.

	Model 1 [17]	Model 2 [16]	Model 3 [42]	Proposed model
IQ metrics / features	Authority/Reputation Completeness Complexity Informativeness Consistency Currency Volatility	Completeness Informativeness ArticleLength NumHeadings NumRefs/Length	Structural features Readability scores	Expertise Authority Trustworthiness
Accuracy	73%	68%	83%	95%
F1 score	72%	59%	80%	94%

TABLE 5. Comparison of the Proposed Model's Clustering Performance with Model1, Model 2 and Model 3.

Classification Type	Model 1 [17]	Model 2 [16]	Model 3 [42]	Proposed model
FA correctly classified	31.4%	30.3%	39.6%	49%
FA incorrectly classified	15.6%	18%	16.7%	6.4%
Non-FA correctly classified	34.4%	32%	33.3%	43.6%
Non-FA incorrectly classified	18.6%	19.7%	10.4%	1%
Total	2000			

the state-of-the-art approaches. According to the results, the accuracies of Model 1, Model 2 and Model 3 were 73%, 68% and 83%, respectively, along with F1 score of 72%, 59% and 80%, respectively (refer to Table 4). The results show that the suggested model has notably improved classification performance. Hence, the proposed model has resulted in a finer IQ assessment for Wikipedia than most of the existing approaches. Furthermore, the model was tested with four other machine learning algorithms: Logistic Regression, Support Vector Machine (SVM), Naïve Bayes, and KNN with cross-validation and received an accuracy of 96%, 95%, 95% and 94% respectively. F1 scores were 94%, 96%, 92% and 94% respectively.

B. CLUSTERING PERFORMANCE

The model's clustering performance was evaluated as the second evaluation technique. K-means clustering was used by Stvilia *et al.* [17] to assess the power of their IQ measures in terms of how successfully they classified the articles into the correct classes. Therefore, the authors also decided to adopt the same clustering algorithm. Based on the derived values for the 1st and 2nd order IQ metrics, a single numeric IQ value per article was derived using equation (1). This IQ was then clustered using the K-means clustering algorithm and the IQ classes to cluster comparisons were observed. Accordingly, 93% of the articles have

been properly clustered by the proposed model. The clustering was replicated with Models 1, 2 and 3, and the comparisons of the results with the proposed model are given in Table 5.

Accordingly, Models 1, 2 and 3 have appropriately clustered 66%, 62% and 73% of the articles. Therefore, it indicates that the proposed model gives far more insight into Wikipedia's IQ evaluation than earlier models.

C. INTER-RATER RELIABILITY TEST RESULT

The authors adopted Fleiss-Davies Kappa [70] as the third evaluation technique. It measures the level of agreement (inter-rater reliability) between a set of raters. A set of three raters who possess prior knowledge of IQ and Wikipedia were chosen for the process. Firstly, each of them was given the 45 IQ attributes and the set of IQ dimensions to group each of the attributes into one of the IQ dimensions, resulting in a Fleiss-Davies kappa of 80%. Secondly, they were asked to group the set of IQ dimensions into the three constructs: Expertise, Authority and Trustworthiness, resulting in a Fleiss-Davies kappa value of 87%. According to authors in [73], the degree of agreement between raters is represented by Fleiss-Davies kappa. Fleiss-Davies kappa values greater than 0.75 indicate strong agreement above and beyond chance, while less than 0.40 indicates poor agreement above and

beyond chance. Fleiss-Davies kappa values between 0.40 and 0.75 suggest a fair to a good level of agreement. Accordingly, both groups achieved an excellent agreement among raters, with an average Kappa value of 84%, concluding that the proposed model is valid.

VI. DISCUSSION

A. FINDINGS OF THE STUDY

Through this study, several important findings need to be discussed. Firstly, the study introduces a novel framework based on the E-A-T model. The model was evaluated in three ways. It scored a 95% of classification accuracy, a 93% of clustering accuracy, and an 84% of Fleiss-Davies Kappa value. Accordingly, the study's findings reveal that the proposed model delivers a better Wikipedia IQ assessment than earlier models. This implies that the E-A-T model can be adopted to assess Wikipedia's IQ. Through this model, the authors propose Expertise, Authority and Trustworthiness as the novel and the most precise constructs of IQ for the Wikipedia encyclopedia. The regression analysis proved that all three constructs improve Wikipedia's IQ.

Secondly, a set of IQ dimensions that influence the above three IQ constructs was presented. Accordingly, the Expertise can be expressed in terms of information richness (informativeness), how well the reader can understand the presented information (understandability) and what amount of effort the reader should exert to read a particular piece of information (readability). Authority also improves the IQ, which means that when other experts cite the content, it implies that the relevant Wikipedia articles are of high quality. Higher Trustworthiness also leads to higher IQ. Trustworthiness can be expressed in the ability to rely on a particular article and how well that article maintains its quality over time (reliability), state of being capable of confirmed (verifiability), and maturity of the content.

Thirdly, a comprehensive set of 45 IQ attributes was presented. These 45 attributes were used to measure each of the above IQ dimensions. Fourthly, an efficient web scraping mechanism was employed to automatically extract the IQ attributes from Wikipedia content and meta-data to overcome the limitations in previous studies' manual feature extraction approaches [42].

B. IMPLICATIONS

There are several theoretical implications of this study. The study adds to the body of knowledge a novel IQ assessment model for Wikipedia based on Google's E-A-T model, which comprises (1) three precise IQ constructs Expertise, Authority and Trustworthiness (2) a set of IQ dimensions that influence the above constructs and (3) 45 IQ attributes to measure the above dimensions.

In terms of practical implications, the findings of this study assist readers in identifying high-quality Wikipedia articles. Thus, rather than labelling it as an unreliable source of information in the big picture, this approach can be presented to

extract high quality and credible information from Wikipedia, the largest source of freely available online information store on the web. On the other hand, based on the proposed model, the study suggests an automated approach to identifying the flaws of articles that would provide immediate guidance for reviewers, authors and editors to implement the quality improvements in the Wikipedia community, which consumes much time and human effort. Furthermore, it is important to note that this approach is not a replacement for the existing manual reviewing method but should be used as an initial screening tool and a supportive tool during the formal review process.

C. LIMITATIONS AND FURTHER WORK

Firstly, the authors tried to generalize the study results by collecting the data from six different domains: Medicine, Politics, Sports, History, Science and Biographies. However, repeating the study for other domains and increasing the size of the dataset would make the results more generalized than this. Secondly, this study focuses on English language Wikipedia articles only. Wikipedia articles, on the other hand, are available in over 300 different languages. Thus, similar studies in other languages also can be conducted. Furthermore, the presented E-A-T based model with three constructs and the set of IQ dimensions can also be generalized and adapted to assess the IQ of other online collaborative repositories and User Generated Content (UGC) by replacing them with new IQ attributes available in those repositories followed through the hybrid approach.

VII. CONCLUSION

In conclusion, this study provided a new perception of adopting the E-A-T model to assess the Wikipedia article's IQ. It was proven and validated that this model provided an improved IQ assessment for Wikipedia than the existing methods. Assessing the IQ of Wikipedia benefits both readers and collaborators. It directs readers toward high-quality articles and guides collaborators to identify the flaws of the articles and plan for further quality improvements. Furthermore, based on the empirical findings and current analysis, it is worth noting that the hybrid approach of feature extraction which is a blend of content and meta-data statistics aids in the construction and evaluation of IQ variance of any online collaborative resource in a cost-effective and scalable manner. The authors intend to extend this study by improving the dataset by size and domain. Furthermore, extending the model to other commonly used languages of Wikipedia articles is also listed as a future work of this study. On the whole, the findings of this study provide good insights towards assessing the IQ of Wikipedia articles and efficient knowledge collaboration in the Wikipedia community.

APPENDIX A

See Table 6.

TABLE 6. Information Quality Attributes Extracted from Article Content and Meta-data Statistics.

No	Information Quality measurement	Description	Source
1	Info Noise	This is calculated as, $1 - (\text{The size of the term/token vector after stemming and stopping}) / (\text{document size before processing})$ [1]	Article
2	Total Num. Edits	Number of edits in an article [1]	Meta-data
3	Num. Unique Editors	Number of distinct editors per article [1]	Meta-data
4	Diversity	This is calculated using, $\text{Num. Unique Editors} / \text{Total Num. Edits}$ [1]	Meta-data
5	Num. Images	Number of images placed within the text of the article [1]	Article
6	Num. Internal Broken Links	Number of links whose destinations are broken or unavailable [1]	Article
7	Num. Internal Links	Number of links from one Wikipedia article to another page in the Wikipedia Community itself [1]	Article
8	Num. Reverted Edits	Number of edits that have been reverted original/ previous version [1]	Meta-data
9	Median Revert Time	In minutes [1]	Meta-data
10	Age	Article age in days [1]	Meta-data
11	Currency	The time between the dump date and the date of the last update of the article in days [1]	Meta-data
12	Connectivity	Number of Articles connected to a particular article through common editors [1]	Article
13	Num. External Links	Number of links from a Wikipedia page to external sources outside Wikipedia [1]	Article
14	Num. Registered User Edits	Number of edits of a page made by registered users [1]	Meta-data
15	Num. Anonymous User Edits	Number of edits of a page made by anonymous users [1]	Meta-data
16	Flesch Kincaid Grade Level	This score is a modification of the Flesch reading score [2], which estimates the difficulty of reading and comprehending an article [3]	Article
17	Gunning Fog Score	This score can measure the simplicity of the text, which is correlated with the reading level [4]	Article
18	SMOG Index	This score estimates the reading grade that readers had to attain to interpret the text [5]	Article
19	Coleman Liau Index	This index estimates the reading level needed to understand an article [6]	Article
20	Automated Readability Index	This indicator approximates the age needed to understand an article [7] [8]	Article
21	Total Num. Headings	Number of headings in a Wikipedia page [9]	Article
22	Num. Headings (Level 1)	Number of Level-1 subheadings [9]	Article
23	Num. Headings (Level 2)	Number of Level-2 subheadings [9]	Article
24	Page Authority	This measure indicates the authority of the content of a single Wikipedia page over the web. This is measured by how well a specific page is ranked in SERP (Search Engine Results Page). This is has a logarithmic scale that ranges from 1 to 100 [10]	Article
25	Num. Linking Domains	This score measures the number of unique external links to a website. In case of a website has more than one link from the same website, it is considered as one linking root domain [10]	Article
26	Num. Inbound Links	The inbound links/ backlinks refer to the hyperlinks from another web page to a page on the particular website [10]	Article
27	Num. Ranking Keywords	Number of keywords of a particular website includes among the top 50 keywords used by Google's ranking [10]	Article
28	Num. Words	Number of words of an article [11]	Article
29	Num. Sentences	Number of sentences in an article [11]	Article
30	Num. Complex Words	Number of complex words in an article [11]	Article
31	Complex Words Percentage	Percentage of complex words over the total number of words in an article [11]	Article

TABLE 6. (Continued.) Information Quality Attributes Extracted from Article Content and Meta-data Statistics.

32	Average Words Per Sentence	Average words per sentence in an article [11]	Article
33	Average Syllables Per Word	Average syllables per word in an article [11]	Article
43	Minor Edits	Number of edits that the editor believes require no review and could never be the subject of a dispute [12]	Meta-data
35	Bot Edits	Number of edits made by automated software tools for a page [13]	Meta-data
36	Semi-Automated Edits	Number of edits made by semi-automated software tools for a page [14]	Meta-data
37	Edits Made By Top 10% Editors	Number of edits made by the set of editors ranked among the top 10% of the editors [15]	Meta-data
38	Page Size	This measure is calculated bytes, including the entire page's size, including text, images, maps, etc. [16]	Article
39	Num. Copyrighted Images	Certain images in Wikipedia are copyrighted, while some images are not. An image with copyright status is an indication of IQ [17]. Thus, this measure indicates the number of copyrighted images in an article.	Article
40	Num. Characters	Number of all the characters (numbers, letters and symbols, etc.) included on a page [9]	Article
41	Num. References	Number of references included in an article [9]	Article
42	Num. Page Watchers	For any particular page, it is possible to discover how many users have it on their watchlist. When a page is included in one's watch list, he will be able to track any changes done to that page [18]	Meta-data
43	Num. Page Views in 60 Days	Number of views for a page within the last 60 days [19]	Meta-data
44	Num. Edits Made In Past 30 Days	Number of edits made to a page within the last 30 days [20]	Meta-data
45	Average Time Between Edits	Average time between edits of the page [20]	Meta-data

Note - The dataset was collected as of 25/09/2021.

- [1] B. Stvilia, M. B. Twidale, L. C. Smith, and L. Gasser, "Assessing information quality of a community-based encyclopedia," in *Proc. Int. Conf. Inf. Quality (ICIQ)*, 2005, pp. 442–454.
- [2] R. Flesch, "A new readability yardstick," *J. Appl. Psychol.*, vol. 32, no. 3, pp. 221–233, Jun. 1948, doi: 10.1037/h0057532.
- [3] J. Kincaid, R. Fishburne, R. Rogers, and B. Chissom. (Jan. 1975). *Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel*. Inst. Simul. Train. Accessed: Jun. 9, 2021. [Online]. Available: <https://stars.library.ucf.edu/istlibrary/56>
- [4] R. Gunning, "The fog index after twenty years," *J. Bus. Commun.*, vol. 6, no. 2, pp. 3–13, Jan. 1969, doi: 10.1177/002194366900600202.
- [5] G. H. McLaughlin, "SMOG grading: A new readability formula," *J. Read.*, vol. 12, no. 8, pp. 639–646, 1969. Accessed: Jun. 9, 2021. [Online]. Available: <https://psycnet.apa.org/record/1969-14260-001>.
- [6] M. Coleman and T. L. Liau, "A computer readability formula designed for machine scoring," *J. Appl. Psychol.*, vol. 60, no. 2, pp. 283–284, Apr. 1975, doi: 10.1037/h0076540.
- [7] M. A. Suryanto, E. P. Lim, A. Sun, and R. H. L. Chiang, "Quality-aware collaborative question answering: Methods and evaluation," in *Proc. 2nd ACM Int. Conf. Web Search Data Mining (WSDM)*, Barcelona, Spain, 2009, pp. 142–151, doi: 10.1145/1498759.1498820.
- [8] R. J. Senter and E. A. Smith. (Nov. 1, 1967). *Automated Readability Index*. Accessed: Jun. 7, 2021. [Online]. Available: <https://apps.dtic.mil/sti/citations/AD0667273>
- [9] A. Shen, J. Qi, and T. Baldwin, "A hybrid model for quality assessment of Wikipedia articles," in *Proc. Australas. Lang. Technol. Assoc. Workshop*, 2017, pp. 43–52.
- [10] *Learn SEO—Free SEO Learning Center [2021]—Moz*. Accessed: Feb. 26, 2021. [Online]. Available: <https://moz.com/learn/seo>
- [11] (2022). *Readability Test | WebFX*. Accessed: Jan. 1, 2022. [Online]. Available: <https://www.webfx.com/tools/read-able>
- [12] (2022). *Help: Minor Edit—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Help:Minor_edit
- [13] (2022). *Wikipedia: Bots—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: <https://en.wikipedia.org/wiki/Wikipedia:Bots>
- [14] (2022). *Automated Edits—OpenStreetMap Wiki*. Accessed: Jan. 1, 2022. [Online]. Available: https://wiki.openstreetmap.org/wiki/Automated_edits
- [15] (2022). *Edit Count—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Edit_count
- [16] (2022). *Help: Page Size—Meta*. Accessed: Jan. 1, 2022. [Online]. Available: https://meta.wikimedia.org/wiki/Help:Page_size
- [17] Q. V. Dang and C.-L. Ignat, "Quality assessment of Wikipedia articles without feature engineering," in *Proc. 16th ACM/IEEE-CS Joint Conf. Digit. Libraries*, New York, NY, USA, Jun. 2016, pp. 27–30.
- [18] (2022). *Help: Watchlist—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: <https://en.wikipedia.org/wiki/Help:Watchlist#watchers>
- [19] (2022). *Wikipedia: Pageview statistics—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Wikipedia:Pageview_statistics
- [20] (2022). *Help: Editing—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: <https://en.wikipedia.org/wiki/Help:Editing>

REFERENCES

- [1] (2022). *Wikipedia: Size of Wikipedia—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia

- [2] A. G. West, J. Chang, K. Venkatasubramanian, O. Sokolsky, and I. Lee, "Link spamming Wikipedia for profit," in *Proc. 8th Annu. Collaboration, Electron. Messaging, Anti-Abuse Spam Conf.*, vol. 2011, pp. 152–161, doi: 10.1145/2030376.2030394.
- [3] (2022). *Vandalism on Wikipedia—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Vandalism_on_Wikipedia
- [4] J. R. Martínez-Rico, J. Martínez-Romo, and L. Araujo, "Can deep learning techniques improve classification performance of vandalism detection in wikipedia?" *Eng. Appl. Artif. Intell.*, vol. 78, pp. 248–259, Feb. 2019, doi: 10.1016/j.engappai.2018.11.012.
- [5] B. Thomas Adler, L. de Alfaro, S. M. Mola-Velasco, P. Rosso, and A. G. West, "Wikipedia vandalism detection: Combining natural language, metadata, and reputation features," in *Proc. Int. Conf. Intell. Text Process. Comput. Linguistics*, vol. 6609, 2011, pp. 277–288, doi: 10.1007/978-3-642-19437-5_23.
- [6] A. Sarabadani, A. Halfaker, and D. Taraborelli, "Building automated vandalism detection tools for Wikidata," in *Proc. 26th Int. Conf. World Wide Web Companion (WWW Companion)*, 2017, pp. 1647–1654, doi: 10.1145/3041021.3053366.
- [7] H. Meishar-Tal, "Teachers' use of Wikipedia with their students," *Austral. J. Teacher Edu.*, vol. 40, no. 12, p. 40, Jan. 2015, doi: 10.14221/ajte.2015v40n12.9.
- [8] *School Officials Unite in Banning Wikipedia | The Seattle Times*. Accessed: May 4, 2021. [Online]. Available: <https://www.seattletimes.com/life/lifestyle/school-officials-unite-in-banning-wikipedia/>
- [9] P. Konieczny, "Rethinking Wikipedia for the classroom," *Contexts*, vol. 13, no. 1, pp. 80–83, Feb. 2014, doi: 10.1177/1536504214522017.
- [10] J. A. Hoffer, M. B. Prescott, and F. R. McFadden, *Modern Database Management*. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.
- [11] M. Gustavsson and C. Wänström, "Assessing information quality in manufacturing planning and control processes," *Int. J. Quality Rel. Manage.*, vol. 26, no. 4, pp. 325–340, Apr. 2009, doi: 10.1108/02656710910950333.
- [12] K. P. Fadahunsi, J. T. Akinlua, S. O'Connor, P. A. Wark, J. Gallagher, C. Carroll, A. Majeed, and J. O'Donoghue, "Protocol for a systematic review and qualitative synthesis of information quality frameworks in eHealth," *BMJ Open*, vol. 9, no. 3, Mar. 2019, Art. no. e024722, doi: 10.1136/BMJOPEN-2018-024722.
- [13] *Characteristics of an Encyclopedia*. Britannica. Accessed: Aug. 27, 2021. [Online]. Available: <https://www.britannica.com/summary/encyclopaedia>
- [14] (2022). *Wikipedia: Statistics—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: <https://en.wikipedia.org/wiki/Wikipedia:Statistics>
- [15] (2022). *Wikipedia: Content Assessment—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Wikipedia:Content_assessment
- [16] M. Warncke-Wang, D. Cosley, and J. Riedl, "Tell me more: An actionable quality model for Wikipedia," in *Proc. 9th Int. Symp. Open Collaboration*, Aug. 2013, pp. 1–10, doi: 10.1145/2491055.2491063.
- [17] B. Stvilia, M. B. Twidale, L. C. Smith, and L. Gasser, "Assessing information quality of a community-based encyclopedia," in *Proc. Int. Conf. Inf. Quality (ICIQ)*, 2005, pp. 442–454.
- [18] M. Anderka, B. Stein, and N. Lipka, "Predicting quality flaws in user-generated content: The case of Wikipedia," in *Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, 2012, pp. 981–990, doi: 10.1145/2348283.2348413.
- [19] *Search Quality Evaluator Guidelines*, Google, Mountain View, CA, USA, 2018.
- [20] A. Kittur and R. E. Kraut, "Harnessing the wisdom of crowds in wikipedia: Quality through coordination," in *Proc. ACM Conf. Comput. Supported Cooperat. Work (CSCW)*, 2008, pp. 37–46, doi: 10.1145/1460563.1460572.
- [21] D. M. Wilkinson and B. A. Huberman, "Assessing the value of cooperation in Wikipedia," *1st Monday*, vol. 12, no. 4, Apr. 2007. [Online]. Available: http://firstmonday.org/issues/issue12_4/wilkinson/index.html, doi: 10.5210/fm.v12i4.1763.
- [22] O. Arazy and O. Nov, "Determinants of Wikipedia quality: The roles of global and local contribution inequality," in *Proc. ACM Conf. Comput. Supported Cooperat. Work (CSCW)*, 2010, pp. 233–236, doi: 10.1145/1718918.1718963.
- [23] K. Stein and C. Hess, "Does it matter who contributes: A study on featured articles in the German Wikipedia," in *Proc. 18th Conf. Hypertext Hypermedia*, New York, NY, USA, 2007, pp. 171–174, doi: 10.1145/1286240.1286290.
- [24] K. Nemoto, P. Gloor, and R. Laubacher, "Social capital increases efficiency of collaboration among Wikipedia editors," in *Proc. 22nd ACM Conf. Hypertext Hypermedia (HT)*, 2011, pp. 231–240, doi: 10.1145/1995966.1995997.
- [25] J. E. Blumenstock, "Size matters: Word count as a measure of quality on Wikipedia," in *Proc. Proc. 17th Int. Conf. World Wide Web (WWW)*, Beijing, China, 2008, pp. 1095–1096, doi: 10.1145/1367497.1367673.
- [26] M. A. Suryanto, E. P. Lim, A. Sun, and R. H. L. Chiang, "Quality-aware collaborative question answering: Methods and evaluation," in *Proc. 2nd ACM Int. Conf. Web Search Data Mining (WSDM)*, Barcelona, Spain, 2009, pp. 142–151, doi: 10.1145/1498759.1498820.
- [27] N. Lipka and B. Stein, "Identifying featured articles in Wikipedia: Writing style matters," in *Proc. 19th Int. Conf. World Wide Web (WWW)*, Raleigh, NC, USA, 2010, pp. 1147–1148, doi: 10.1145/1772690.1772847.
- [28] H. Zhang, Y. Ren, and R. Kraut, "Mining and predicting temporal patterns in the quality evolution of Wikipedia articles," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, 2020, pp. 1–10, doi: 10.24251/hicss.2020.485.
- [29] D. H. Dalip, M. A. Gonçalves, M. Cristo, and P. Calado, "Automatic quality assessment of content created collaboratively by web communities: A case study of Wikipedia," in *Proc. 13th Int. Symp. Open Collaboration*, Austin, TX, USA, 2009, pp. 295–304, doi: 10.1145/1555400.1555449.
- [30] P. Wang and X. Li, "Assessing the quality of information on Wikipedia: A deep-learning approach," *J. Assoc. Inf. Sci. Technol.*, vol. 71, no. 1, pp. 16–28, Jan. 2020, doi: 10.1002/asi.24210.
- [31] B. P. R. Guda, S. B. Seelaboyina, S. Sarkar, and A. Mukherjee, "NwQM: A neural quality assessment framework for Wikipedia," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 8396–8406, doi: 10.18653/v1/2020.emnlp-main.674.
- [32] Q. V. Dang and C. L. Ignat, "An end-to-end learning solution for assessing the quality of Wikipedia articles," in *Proc. 13th Int. Symp. Open Collaboration*, vol. 2017, pp. 1–10, doi: 10.1145/3125433.3125448.
- [33] K. D. C. G. Kapugama, S. A. S. Lorensuhewa, and M. A. L. Kalyani, "Enhancing Wikipedia search results using text mining," in *Proc. 16th Int. Conf. Adv. ICT Emerg. Regions (ICTer)*, Sep. 2016, pp. 168–175, doi: 10.1109/ICTER.2016.7829915.
- [34] A. Levitin and T. Redman, "Quality dimensions of a conceptual view," *Inf. Process. Manage.*, vol. 31, no. 1, pp. 81–88, Jan. 1995, doi: 10.1016/0306-4573(95)80008-H.
- [35] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *J. Manage. Inf. Syst.*, vol. 12, no. 4, pp. 5–33, 1996.
- [36] M. Hu, E.-P. Lim, A. Sun, H. W. Lauw, and B.-Q. Vuong, "Measuring article quality in Wikipedia: Models and evaluation," in *Proc. 16th ACM Conf. Conf. Inf. Knowl. Manage. (CIKM)*, 2007, pp. 243–252, doi: 10.1145/1321440.1321476.
- [37] J. Liu and S. Ram, "Who does what: Collaboration patterns in the Wikipedia and their impact on article quality," *ACM Trans. Manage. Inf. Syst.*, vol. 2, no. 2, pp. 1–23, Jun. 2011, doi: 10.1145/1985347.1985352.
- [38] M. A. Sicilia, N. T. Korfiatis, M. Poulos, and G. Bokus, "Evaluating authoritative sources using social networks: An insight from Wikipedia," *Online Inf. Rev.*, vol. 30, no. 3, pp. 252–262, 2006, doi: 10.1108/14684520610675780.
- [39] G. Wu, M. Harrigan, and P. Cunningham, "Classifying Wikipedia articles using network motif counts and ratios," in *Proc. 8th Annu. Int. Symp. Wikis Open Collaboration (WikiSym)*, 2012, pp. 1–10, doi: 10.1145/2462932.2462948.
- [40] T. Wöhner and R. Peters, "Assessing the quality of Wikipedia articles with lifecycle based metrics," in *Proc. 5th Int. Symp. Wikis Open Collaboration (WikiSym)*, 2009, doi: 10.1145/1641309.1641333.
- [41] M. Warncke-Wang, V. R. Ayukae, B. Hecht, and L. Terveen, "The success and failure of quality improvement projects in peer production communities," in *Proc. 18th ACM Conf. Comput. Supported Cooperat. Work Social Comput.*, Feb. 2015, pp. 743–756, doi: 10.1145/2675133.2675241.
- [42] A. Shen, J. Qi, and T. Baldwin, "A hybrid model for quality assessment of Wikipedia articles," in *Proc. Australas. Lang. Technol. Assoc. Workshop, Brisbane, QLD, Australia*, 2017, pp. 43–52.
- [43] W. Lewoniewski, K. Węcel, and W. Abramowicz, "Relative quality and popularity evaluation of multilingual Wikipedia articles," *Informatics*, vol. 4, no. 4, p. 43, 2017, doi: 10.20944/preprints201709.0130.v1.
- [44] D. Jemielniak and M. Wilamowski, "Cultural diversity of quality of information on wikipeidias," *J. Assoc. Inf. Sci. Technol.*, vol. 68, no. 10, pp. 2460–2470, Oct. 2017, doi: 10.1002/ASL.23901.
- [45] C. Sugandhika, S. Ahangama, and S. Ahangama, "Modelling Wikipedia's information quality using informativeness, reliability and authority," in *Proc. 3rd Int. Conf. Advancements Comput. (ICAC)*, Dec. 2021, pp. 169–174, doi: 10.1109/ICAC54203.2021.9671092.

- [46] Q.-V. Dang and C.-L. Ignat, "Measuring quality of collaboratively edited documents: The case of Wikipedia," in *Proc. IEEE 2nd Int. Conf. Collaboration Internet Comput. (CIC)*, Nov. 2016, pp. 266–275, doi: [10.1109/CIC.2016.044](https://doi.org/10.1109/CIC.2016.044).
- [47] Q. V. Dang and C.-L. Ignat, "Quality assessment of Wikipedia articles without feature engineering," in *Proc. 16th ACM/IEEE-CS Joint Conf. Digit. Libraries*, Newark, NJ, USA, Jun. 2016, pp. 27–30.
- [48] Y. Javed, K. M. Salehin, and M. Shehab, "A study of south Asian websites on privacy compliance," *IEEE Access*, vol. 8, pp. 156067–156083, 2020, doi: [10.1109/ACCESS.2020.3019334](https://doi.org/10.1109/ACCESS.2020.3019334).
- [49] I. Booth, *Google EAT: How to Create Content That Google Wants—Moz*. Accessed: Feb. 26, 2021. [Online]. Available: <https://moz.com/blog/google-e-a-t>
- [50] Y. Suzuki, "Quality assessment of Wikipedia articles using h-index," *J. Inf. Process.*, vol. 23, no. 1, pp. 22–30, 2015, doi: [10.2197/IPSJIP.23.22](https://doi.org/10.2197/IPSJIP.23.22).
- [51] X. Li, J. Tang, T. Wang, Z. Luo, and M. de Rijke, "Automatically assessing Wikipedia article quality by exploiting article–editor networks," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes Artificial Intelligent Lecture Notes Bioinformatics)*, vol. 9022. Cham, Switzerland: Springer, 2015, pp. 574–580, doi: [10.1007/978-3-319-16354-3_64](https://doi.org/10.1007/978-3-319-16354-3_64).
- [52] Y. Xu and T. Luo, "Measuring article quality in Wikipedia: Lexical clue model," in *Proc. 3rd Symp. Web Soc.*, Oct. 2011, pp. 141–146, doi: [10.1109/SWS.2011.6101286](https://doi.org/10.1109/SWS.2011.6101286).
- [53] L. R. Jimenez, M. Solera, M. Toril, C. Gijon, and P. Casas, "Content matters: Clustering web pages for QoE analysis with WebCLUST," *IEEE Access*, vol. 9, pp. 123873–123888, 2021, doi: [10.1109/ACCESS.2021.3110370](https://doi.org/10.1109/ACCESS.2021.3110370).
- [54] W. Lewoniewski, K. Węcel, and W. Abramowicz, "Multilingual ranking of Wikipedia articles with quality and popularity assessment in different topics," *Computers*, vol. 8, no. 3, p. 60, Aug. 2019, doi: [10.3390/COMPUTERS8030060](https://doi.org/10.3390/COMPUTERS8030060).
- [55] W. Lewoniewski, K. Węcel, and W. Abramowicz, "Quality and importance of Wikipedia articles in different languages," in *Information and Software Technologies (Communications in Computer and Information Science)*, vol. 639. Cham, Switzerland: Springer, 2016, pp. 613–624, doi: [10.1007/978-3-319-46254-7_50](https://doi.org/10.1007/978-3-319-46254-7_50).
- [56] D. Lewandowski and U. Spree, "Ranking of Wikipedia articles in search engines revisited: Fair ranking for reasonable quality?" *J. Amer. Soc. Inf. Sci. Technol.*, vol. 62, no. 1, pp. 117–132, Jan. 2011, doi: [10.1002/ASI.21423](https://doi.org/10.1002/ASI.21423).
- [57] H. Zeng, M. A. Alhossaini, L. Ding, R. Fikes, and D. L. McGuinness, "Computing trust from revision history," in *Proc. Int. Conf. Privacy, Secur. Trust Bridge Gap Between PST Technol. Bus. Services (PST)*, 2006, p. 1, doi: [10.1145/1501434.1501445](https://doi.org/10.1145/1501434.1501445).
- [58] J. M. Nicholson, A. Uppala, M. Sieber, P. Grabitz, M. Mordaunt, and S. Rife, "Measuring the quality of scientific references in Wikipedia: An analysis of more than 115M citations to over 800 000 scientific articles," *FEBS J.*, vol. 288, no. 14, pp. 4242–4248, 2020, doi: [10.1101/2020.04.08.031765](https://doi.org/10.1101/2020.04.08.031765).
- [59] D. Hasan Dalip, M. André Gonçalves, M. Cristo, and P. Calado, "Automatic quality assessment of content created collaboratively by web communities: A case study of Wikipedia," in *Proc. Joint Int. Conf. Digit. Libraries (JC DL)*, 2009, pp. 295–304, doi: [10.1145/1555400.1555449](https://doi.org/10.1145/1555400.1555449).
- [60] C. Sugandhika, S. Ahangama, and S. Ahangama, "Assessing the information quality of Wikipedia articles using web presence," in *Proc. Int. Conf. Innov. Emerg. Technol.*, 2021, p. 37.
- [61] P. Wang and X. Li, "Assessing the quality of information on Wikipedia: A deep-learning approach," *J. Assoc. Inf. Sci. Technol.*, vol. 71, no. 1, pp. 16–28, Jan. 2020, doi: [10.1002/ASI.24210](https://doi.org/10.1002/ASI.24210).
- [62] *Selenium*. Accessed: Sep. 24, 2021. [Online]. Available: <https://www.selenium.dev/>
- [63] (2022). *Wikipedia: WikiProject Medicine—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Medicine
- [64] (2022). *Wikipedia: WikiProject Politics—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Politics
- [65] (2022). *Wikipedia: WikiProject Sports—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Sports
- [66] (2022). *Wikipedia: WikiProject History—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_History
- [67] (2022). *Wikipedia: WikiProject Science—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Science
- [68] (2022). *Wikipedia: WikiProject Biography/Assessment—Wikipedia*. Accessed: Jan. 1, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Biography/Assessment
- [69] D. Child, *The Essentials of Factor Analysis*, 2nd ed. London, U.K.: Cassell Educational, 1990.
- [70] M. Banerjee, M. Capozzoli, L. McSweeney, and D. Sinha, "Beyond kappa: A review of interrater agreement measures," *Can. J. Statist.*, vol. 27, no. 1, pp. 3–23, 1999, doi: [10.2307/3315487](https://doi.org/10.2307/3315487).
- [71] B. J. Babin and R. E. Anderson, *On Multivariate Data Analysis*, J. F. Hair, Jr., and W. C. Black, Eds., 7th ed. Harlow, U.K.: Pearson, 2014.
- [72] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 4th ed. New Delhi, India: SAGE, 2009.
- [73] S. Simon, A. Carbone, M. de Raadt, R. Lister, M. Hamilton, and J. Sheard, "Classifying computing education papers: Process and results," in *Proc. 4th Int. Workshop Comput. Educ. Res. (ICER)*, 2008, pp. 161–171, doi: [10.1145/1404520.1404536](https://doi.org/10.1145/1404520.1404536).



CHINTHANI SUGANDHIKA is currently a Research Scholar with the University of Moratuwa, Sri Lanka. She is a published author at reputed international conferences and has won many awards for her outstanding publications. Her current research interests include data science, natural language processing, software engineering, and cyber security.



SUPUNMALI AHANGAMA (Member, IEEE) received the Ph.D. degree in information systems from the National University of Singapore. She is currently a Senior Lecturer with the Department of Information Technology, University of Moratuwa, Sri Lanka. She is also the Director of Undergraduate Studies with the Faculty of Information Technology, University of Moratuwa. She has presented her work and served as a Reviewer for numerous top-tier forums, including *e-Service Journal*, *IEEE Access*, *Information Systems Frontiers*, International Conference on Information Systems (ICIS), Pacific Asia Conference on Information Systems (PACIS), the Workshop on Information Technologies and Systems (WITS), and HCI International. Her research interests include data science, design science, information systems, and e-education.

• • •