Cultural Determinants of Sharing High-quality Knowledge: A Case Study of Wikipedia's Featured Articles

Katarzyna Kukowska
Czestochowa University of Technology, Poland

katarzyna.kukowska@pcz.pl

Abstract: Wikipedia is the largest encyclopedia in the world. It works by involving users in sharing knowledge, systematising it and correcting the content of encyclopedia entries. The range of features that involve users in the process of creating the encyclopedia and its derivative projects is much wider. From a research perspective, the socio-cultural foundations of this encyclopedia are the most interesting. The aim of this study was to present the ways in which users are involved in creating articles referred to as featured. Respondents participating in the study were asked this question: does the structure and the way authors and editors of high-quality articles are involved in the project remain dependent on the Hofstede Individualism Index (IDV)? The method used in the study was an analysis of user activity and ways in which users were involved. The author obtained the data using the xTools tool available on Wikipedia. When selecting the language versions for the analysis, criteria were used: 1) their cultural homogeneity, 2) extreme IDV values – individualism versus collectivism, in order to obtain a clear contrast, and 3) a large number of featured articles. For this sample, in cultures with high IDV rates, users were involved in creating and modifying featured articles in significantly greater numbers. However, the smallest share of featured articles concerns the language versions with the highest IDV. Statistically, the study showed significant differences in the form of author and editor engagement in creating high-quality articles. The results obtained do provide grounds for concluding that the quality of content depends on Hofstede's cultural indicators. In continuing the study, it is worth considering making a comparison of the number of featured articles of different language versions in particular fields of knowledge.

Keywords: Activity in Virtual Communities, Cultural Dimensions, Sharing Knowledge, Wikipedia Users

1. Introduction

The prospect of creating a sustainable, evolving and effective organisation based on motivated, active and unpaid activists is very attractive cognitively. The Wikipedia project proves the scale on which the cooperation of a huge number of individuals can be organised. Wikipedia does not rely on traditional forms of governance that require hierarchical leadership. The individual's contribution is fully autonomous, based on intrinsic motivation. The absence of user declarations as to the degree of commitment or tools to coerce a formal agreement has not stopped the organisation from working continuously for over 20 years.

Each language version is an autonomous community under the supervision of the Wikimedia Foundation, with Meta-Wiki coordination. Therefore, the socio-cultural determinants of the activity of Wikipedia editors are of interest to the study. The aim of this study was to present the ways in which users were involved in creating articles referred to as featured. Featured articles are a high-quality body of knowledge, created using a process of collective commitment to knowledge sharing. Wikipedians are guided by the following criteria to qualify an article as featured: a comprehensive description of the topic; a text that covers the main aspects of the topic without going into unnecessary details; neutrality; An exhaustively illustrated topic and an exhaustive overview of sources relevant to the topic is used (https://en.wikipedia.org/wiki/Wikipedia:Compare_criteria...). The high quality of the article is confirmed by the award of a symbolic medal or barnstar. Table 1 presents the symbols used to mark featured articles in the language versions selected for the study.

Table 1: Symbols of Featured Articles of Selected Language Versions in Wikipedia

Indonesian	Dutch Thai	Swedish	Georgian Norwegian Slovenian	Persian Vietnamese	German	Czech	Polish
							

Wikipedia exemplifies new patterns of cooperation and new models of information, revision and distribution that are shaping the modern model of authorship (Alonso and Robinson, 2016). The article is a contribution to a study of the determinants of high-quality knowledge sharing. The main question is: does the structure and the way authors and editors of high-quality articles are involved in the project remain dependent on the Hofstede Individualism Index (IDV)?

The article is of exploratory nature. In order to answer the research question the paper consists of sections and namely: section 2 presents a literature review, pointing out Wikipedia's lack of cultural neutrality, section 3 presents description of the research process, section 4 presents study results (hypotheses are verified in the subsections), section 5 provides a discussion and identifies the limitations of the study. The article ends with conclusions.

2. Wikipedia and Cultural Neutrality

Since the inception of Wikipedia, the issue of the quality of content created by Wikipedians has been raised. This is despite the fact that, after only a few years of Wikipedia, Giles (2005) suggested that scholarly articles on Wikipedia had reached a level of quality similar to that of the Encyclopædia Britannica. Nevertheless, Mesgari et al (2015), when identifying Wikipedia-related scholarly publications, showed that about 80% of research queries focused on quality issues. Konieczny (2021) notes that advanced review processes, particularly "Recommended Article Review", provide good protection and oversight over the quality of content. In the case of featured articles, research is largely focused on processes that monitor the quality of content. One factor influencing the quality of articles is the attributes, roles, and experiences of editors (Ren et al, 2016). Articles edited mainly by "versatile" editors, which occur in various Wikipedia activities, tend to be of high quality. In addition, research has linked the quality of articles to the experience of the main co-author (Kane, 2011), the knowledge of editors, and affiliation with WikiProjects (Kittur et al, 2009). Iba et al (2010) have developed a tool that transforms the flow of edits between contributors into a temporary social network. They analysed 2580 articles from English Wikipedia and identified a key category of coolfarmers, authors who start and create new high-quality articles best suited to enforce behaviours consistent with Wikipedia's culture. The quality of Wikipedia is deteriorated by egoboosters, people who use Wikipedia mainly to present themselves by fuelling discussions on controversial topics. Finding indicators to identify Wikipedia's most valuable contributors can be used for alternative article quality ranking systems based on author quality. Zhang et al (2020) analysed archival data of 6057 articles, examining the evolution of the quality of articles on Wikipedia. Some of the articles will never exceed the low or moderate level. Most articles achieve high quality through years of continuous efforts. In addition, the development of the article is influenced by the time the article was created, its popularity and relevance. Based on the results of the quality assessment of more than 39 million articles in 55 languages, Lewoniewski et al (2019) investigated which topics were best represented in different language versions of Wikipedia. It was confirmed that the popularity of an article could be a factor relevant to assessing quality. Articles on topics popular in a given language are of relatively high quality. One important condition is that a small group of editors contributing a disproportionately high percentage of edits remains in order for the text to achieve the status of featured articles. Lewoniewski et al (2016) proposed models to assess the relative content of the entire article and the quality of the data contained in its structural elements. For each language version of Wikipedia, parameters are selected to enable automatic evaluation. Links between articles in different languages provide opportunities to compare and verify the quality of information in different language versions. On the basis of the various characteristics of featured articles, other articles can be evaluated. Lewoniewski et al (2023) also conducted an analysis of Wikipedia references in 310 languages. Additionally, they classified Wikipedia articles according to different topics. This made it possible to find differences in the range of scholarly sources between the language versions of Wikipedia.

Wikipedia is not culturally neutral. Wikipedia editors reflect their own cultural norms by referring to the process of creating Wikipedia articles and different patterns of collaboration. A high IDV stands for individual identity, dominant over group identity. Self-actualisation and a firm private opinion are valued. Collectivism is concerned with building harmony and consensus within a group, avoiding confrontation (Hofstede et al, 2010, pp 89-134). Cultural differences in Wikipedia language versions based on Hofstede dimensions have been identified by Pfeil et al (2006) etc. The authors found that the likelihood of editors performing different types of actions on Wikipedia was correlated with differences in cultural dimensions. Pfeil et al argued that a greater number of activities would be associated with the linguistic version identified with collectivist cultures. A total of 952 changes to the selected entry of the four versions of Wikipedia were analysed. The higher the IDV of a given country, the more entries members of that Wikipedia page contributed in remedial categories. Stvilia et al (2009) measured the cultural similarity between the three Wikipedias using Hofstede's typology. In the study, the authors analysed quality understanding and value structure by tracking voting and decision-making behaviour on featured articles. In the sample, the number of editors and the number of edits made by registered users were negatively related to quality. The relationship between the degree of similarity of cultural traits and the degree of similarity of quality models in Wikipedias was not conclusive. Nemoto and Gloor (2011) analysed the

process of creating articles and social interactions in five language versions of Wikipedia. The authors found differences in communication behaviour between egalitarian and hierarchical cultures.

Despite extensive research, the debate on the quality of content and the resulting quality of knowledge popularised by Wikipedia will continue. As Jemielniak (2019) notes, Wikipedia will remain in the space of social life, providing free encyclopedic knowledge for the good of the public. Therefore, the motivation to study the behaviour of Wikipedia editors in the light of cultural conditions is justified.

3. Methodology

The article presents the next stage of a cyclical stud y on the socio-cultural determinants of the Wikipedia project, which has been in progress since 2021. So far, the author's study has shown, differences in the cooperation structures of the studied language versions of Wikipedia, in relation to selected cultural indicators by Hofstede (Kukowska, 2023; Kukowska, 2022; Skolik, Kukowska, 2021). The starting point for the current research was the involvement of users of selected language versions of Wikipedia in creating and sharing of high-quality knowledge. Co-authorship in featured articles was taken as an indicator of high quality. The level of involvement of featured article creators in various forms of activity is reflected in the individual namespaces. Namespaces are the types of web pages that make up Wikipedia content, and articles are only part of the namespace. Technically, namespaces are a collection of pages whose names begin with a prefix recognised by MediaWiki software (https://en.wikipedia.org/wiki/Wikipedia:Namespace). The study assumes that there is some correlation between the creation of high-quality knowledge by the most active users in this area and cultural conditions. The Individualism index (IDV) in Hofstede's typology was considered as the basis for differences in the cultural dimension. It was assumed that this differentiation is reflected in the form of activities reported in the statistics of the users represented. The following hypotheses were adopted:

H1 – According to the preferred models of work (collective versus individualistic), the featured articles of language versions with high IDV are created by a small group of people, as opposed to cultures with low IDV.

H2 – The index of individualism in a given culture (reflecting the attitude towards cooperation) translates into a higher or lower share of featured articles in particular language versions.

H3 – The distribution of activity of high-quality content creators in namespaces is clearly different for each language version, in view of the IDV value.

A number of steps were taken to carry out the analysis. First of all, from the https://meta.wikimedia.org/wiki/Wikipedia_featured_articles website, language versions with at least 50 featured entries were selected. The IDV indicator was used to qualify the language versions for data analysis. It was assumed that in the process of creating common content by often very many users, it is this cultural dimension that remains crucial to such a working model. The selection of language versions took into account their cultural homogeneity and extreme IDV values, i.e. above 70 or below 30 on a scale from 0 to 100. The analysis was limited to 12 language versions. The exception was the Polish (IDV47) version, which in the process of data analysis in the files was assigned to the version with low IDV. Some language versions of Wikipedia were not examined - those of the most widely used languages in the world (e.g. English, Spanish, French). They were rejected due to differences in the cultures from which their creators came. This applies especially to English, which is widely recognized as an international language. Using the same language does not always mean cultural closeness, so there was a risk of distorting the study results taking into account the level of IDV.

In the second step, 50 featured articles were randomly selected to analyse their statistics, from each of these versions. Contributors were selected from the history of the articles. Out of 600 featured articles, 2904 authors/editors were selected according to the following measures: 1) the highest percentage contribution to the article, up to the tenth position from the list of editors of a given article, 2) exclusion of bots, and 3) contributors of at least 3 featured articles from the pool drawn. 192 users from language versions with low IDV and 179 users with high IDV were deemed characteristic of user activity. The study analysed user activity by obtaining data from Wikipedia using the xTools tool. Data for the study was collected from January to March 2024.

4. Study Results

4.1 Co-Authorship Structure for Featured Articles – Collectivism Versus Individualism (Verifying H1)

To verify H1, it was necessary to answer the question of what is the structure of co-authorship of high-quality articles. The main axis representing the differences in cultural determinants for creating high-quality content was the division of the language versions of Wikipedia into a set of cultures with a high level of collectivism as opposed to individualistic cultures. After creating two sets of selected language versions overall, the statistics showed a significant difference in the number of total users. Language versions with a low IDV were created by 16,124 active users, and versions with a high IDV had 12,702 more active users (the big difference is due to the very large German version). The total number of editors creating the selected featured articles (before the selection) was adequate to the number of active users. In the language versions of collectivist cultures, 300 articles were produced by 18051 users (median: 2827.5). In individualistic cultures, it was 30,843 users (median: 4,643) per 300 articles. For this sample, in cultures with high IDV rates, users were involved in creating and modifying featured articles in significantly greater numbers. H1 received a negative assessment during verification.

4.2 Contribution of Featured Articles to Encyclopedia Entries (Verifying H2)

The answer to the question of what part of all articles are featured articles in the analyzed language versions of Wikipedia made it possible to verify H2. In the analysed language versions, featured articles account for between 1.7% and 22.1% of all articles in total. Therefore, there is a great discrepancy. The smallest share of featured articles concerns the language versions with the highest IDV: Swedish (IDV87) – 1.4% and Dutch (IDV100) – 1.7%. At the same time, Slovenian (IDV81) has the largest share of featured articles – 22.1%. The set with a lower IDV includes Georgian (IDV15) – 14.4% and Thai (IDV19) – 12.5%. The above discrepancy in the collections accepted leads to the idea that the sheer number of featured articles does not have to be represented in terms of a given cultural dimension. Moreover, the average share of featured articles in total articles for language versions with low IDV is 7.35%, and for language versions with high IDV it is 7.48%. Although this is a small sample, this very small difference in the averages is surprising. In the examples studied, the least numerous language versions of Wikipedia have the largest percentage of featured articles in their encyclopedia entries: Georgian – 284 active users (14.4%), Thai – 1169 (12.5%), Slovenian – 389 (22.1%). H2 has not been confirmed.

4.3 Activities of High-Quality Content Contributors and IDV (verifying H3)

The verification of H3 required an answer to the question of what forms of activity the contributors of high-quality content engage in the most. The Wikipedia statistic describes the characteristics of a Wikipedian's activity by counting the number of edits in a set of MediaWiki pages collectively named namespace (https://en.wikipedia.org/wiki/Wikipedia: Namespace). Namespaces are numerous, and their use varies from language to language. In order to represent user activity, namespaces were grouped according to their functions into the following areas: substantive, technical, interaction (social), content promotion, and working (organisational).

1. Substantive spaces: Main/Article, Category, File and Timedtext are types of namespaces containing content created for readers (recipients). From the encyclopedia's point of view, Mains are the most important. Managing the content of articles (this is the constantly negotiated taxonomy of encyclopedia knowledge) is enabled by Categories. The File namespace contains mainly photos but also other graphic and multimedia forms, and a number of different files that are attachments placed in articles. TimedText, on the other hand, is a multimedia file space that enables the user to add text to video recordings.

2. Technical spaces: Template, Module and MediaWiki contain mainly complex code that enables the user to handle other namespaces. Templates can be more or less complex; they contain rather simple content placed on many pages. One example is the Infobox, which is usually placed in the upper right-hand part of the article; it is technically advanced and can be filled in by entering a small amount of data. A module is a more advanced type of page that some templates use. MediaWiki, on the other hand, is the core software architecture of the Wikimedia projects.

3. Interaction/social spaces: User talk, Wikipedia, Wikipedia talk, Help and Help talk are spaces primarily used for communication between Wikipedians and to inform users about the rules of contributing to an encyclopedia. Originally, this was the purpose of the User talk space. However, such communication was not public, which is why the Wikipedia namespace was created. Common rules were also established in this space, and once

established, the Wikipedia talk space enabled them to be criticised or renegotiated. Help and Help talk have been separated for the benefit of how-to websites.

4. Content promotion spaces: Portal, Book and Story are the types of pages that allow Wikipedians to present content from specific subject areas. Portals are used for internal promotion of certain categories of entries and created according to the model of the main page, introducing the reader to a given topic area. Book and Story, on the other hand, enable individual users to choose content that is interesting to them and promote it outside Wikipedia.

5. Workspaces/Organisational Spaces are any pages where various preparatory work is done, cooperation with other users is undertaken, content is reviewed in detail, and articles are drafted. Dedicated spaces for these purposes are: Wikiproject, Wikiproject talk, Draft and Draft talk. In the case of other spaces, they are also used for the same purpose, although their purpose was different. The User space is used for self-presentation, but often users use it (or rather its subspaces) to create a number of drafts. Talk is a space for evaluating an article, and comments on how to improve an encyclopedia entry are often exchanged here. Talk spaces assigned to other namespaces are treated similarly.

In order to represent the forms of activity in which co-creators of high-quality content engage the most, the Mann-Whitney U test was conducted (Szajt, 2014). The test results are shown in Tables 2-6. Statistically significant differences for the given language versions are marked with an arrow symbol. The direction of the arrow indicates the language version in which users are more active. For the remaining cases, no statistically significant differences were confirmed.

In the case of activity in substantive spaces, the test only showed statistically significant differences in the comparison of language versions with a high IDV index, even in relation to each other. The dominant figures in this system are German and Norwegian, in which only in 3 cases were statistically significant differences not confirmed (for Persian, Thai and Dutch). In substantive areas, it is the co-authors of culturally individualistically conditioned language versions that show dominant activity in relation to collectively conditioned language versions. The only exception is the Czech version, which is inferior to the other versions from its pool (Table 2).

Table 2: Mann-Whitney U test for Author and Editor Activity in Content Spaces

LANGUAGE VERSIONS	IDV	id	ka	th	fa	vi	pl	cs	de	no	sl	sv	nl
id-Indonesian	5												
ka-Georgian	15	-											
th-Thai	19	-	-										
fa-Persian	23	-	-	-									
vi-Vietnamese	30	-	-	-	-								
pl-Polish	47	-	-	-	-	-							
cs-Czech	70	-	-	-	-	-	-						
de-German	79	←	←	-	-	←	←	←					
no-Norwegian	81	←	←	-	-	←	←	←	-				
sl-Slovenian	81	-	-	-	-	-	-	-	-	-			
sv-Swedish	87	←	-	-	-	←	-	-	-	-	-		
nl- Dutch	100	←	←	-	-	←	-	←	-	-	-	-	

In technical areas, all language versions selected for the study showed statistically significant differences in distribution levels in the test in at least 2 cases (e.g. Indonesian with German and Polish). In technical spaces, their users are more active. In this group, the Thai version is predominant in 9 cases, 5 of which refer to individualistic cultures. The trend is offset by the Polish language version, which shows statistically significant differences with all language versions from its set (low IDV) and with half of the language versions with high IDV. At the same time, in every case the Polish language version shows lower user activity in technical spaces. However, this version is an exception that does not show extreme values for the indicator (IDV47) in relation to the others. Cases of language versions with high IDV, in which users are more active (Norwegian, Slovenian and Swedish) are still distributed along with the German language version belonging to their set (Table 3).

Table 3: Mann-Whitney U Test for Author and Editor Activity in Technical Spaces

LANGUAGE VERSIONS	IDV	id	ka	th	fa	vi	pl	cs	de	no	sl	sv	nl
id-Indonesian	5												
ka-Georgian	15	-											
th-Thai	19	←	←										
fa-Persian	23	-	-	-									
vi-Vietnamese	30	-	-	↑	-								
pl-Polish	47	↑	↑	↑	↑	↑							
cs-Czech	70	-	↑	↑	↑	↑	-						
de-German	79	↑	↑	↑	↑	↑	-	-					
no-Norwegian	81	-	-	↑	↑	-	←	-	←				
sl-Slovenian	81	-	-	-	-	-	←	-	←	-			
sv-Swedish	87	-	-	↑	↑	-	←	-	←	-	-		
nl- Dutch	100	-	↑	↑	↑	↑	-	-	-	-	-	-	

User activity in interaction spaces did not show such clear tendencies of dominance of one of the sets as in the case of substantive or technical spaces. In the interaction spaces, no statistically significant differences were confirmed for the Swedish language version. In view of the test results, the Swedish version can be treated as a "typical" version. On one hand, in terms of activity it constitutes the threshold to be reached for collectivist cultures and the lower limit for individualistic cultures. For Norwegian and Vietnamese, one case was shown each. On the other hand both versions give way to the more active Persian version which belongs to a collectivist culture. Compared to users representing different Persian cultures, the language version definitely stood out. For the Persians, 8 advantages were shown: over all versions from their own pool (low IDV) and half of the versions assigned to the individualistic culture (Czech, Norwegian, Slovenian). The Persian version is counterbalanced by the Dutch and German versions with 6 and 5 advantages, respectively, but in their own cultural group (Table 4).

Table 4: Mann-Whitney U Test for Authors and Editor Activity in Interaction Spaces

LANGUAGE VERSIONS	IDV	id	ka	th	fa	vi	pl	cs	de	no	sl	sv	nl
id-Indonesian	5												
ka-Georgian	15	-											
th-Thai	19	-	-										
fa-Persian	23	←	←	←									
vi-Vietnamese	30	-	-	-	↑								
pl-Polish	47	-	-	-	↑	-							
cs-Czech	70	-	←	-	↑	-	-						
de-German	79	←	←	←	-	-	←	-					
no-Norwegian	81	-	-	-	↑	-	-	-	-				
sl-Slovenian	81	-	-	-	↑	-	-	-	↑	-			
sv-Swedish	87	-	-	-	-	-	-	-	-	-	-		
nl- Dutch	100	←	←	←	-	-	←	←	-	-	←	-	

A clear advantage of the activity of authors and editors in the promotion spaces was recorded by the Czech and German versions, defeating in 4 cases representatives of collectivist cultures (Indonesian, Georgian, Vietnamese, Polish) and in 2 cases representatives of their own pool (Norwegian, Slovenian). The versions with the highest IDVs (Swedish and Dutch) dominated the Norwegian and Slovenian versions with slightly lower IDVs. The Polish version of 9 statistically significant differences in no case achieved superiority. Only two versions (Thai and Persian) representing collectivist cultures gained an advantage over the individualist versions: Norwegian and Slovenian (Table 5).

Table 5: Mann-Whitney U Test for Authors' and Editors' Activity in Content Promotion Spaces

LANGUAGE VERSIONS	IDV	id	ka	th	fa	vi	pl	cs	de	no	sl	sv	nl
id-Indonesian	5												
ka-Georgian	15	-											
th-Thai	19	-	-										
fa-Persian	23	-	-	-									
vi-Vietnamese	30	-	-	-	-								
pl-Polish	47	↑	↑	↑	↑	↑							
cs-Czech	70	←	←	-	-	←	←						
de-German	79	←	←	-	-	←	←	-					
no-Norwegian	81	-	-	↑	↑	-	-	↑	↑				
sl-Slovenian	81	-	-	↑	↑	-	-	↑	↑	-			
sv-Swedish	87	-	-	-	-	-	←	-	-	←	←		
nl- Dutch	100	-	-	-	-	-	←	-	-	←	←	-	

The test showed predominant user activity in workspaces for individualistic cultures. Statistically significant differences predicted the favourites of users of two language versions: Dutch and German. Users of the Dutch language version had an advantage over collectivist cultures in 2 cases: Georgian and Indonesian. Other cases of predominance of the Dutch version applied to cultures with high IDV: Czech and Slovenian. The German version had the advantage of greater user activity over 3 versions of its own collection (Slovenian, Norwegian and Czech) and 5 from collectivist cultures (Georgian, Indonesian, Vietnamese, Persian, Thai). As an exception, for the Indonesian version, the test showed a statistically significant difference in distributions compared to all studied language versions except Slovenian. At the same time, the Indonesian version did not gain an advantage in any case (Table 6).

Table 6: Mann-Whitney U Test for Author and Editor Activity in Workspaces

LANGUAGE VERSIONS	IDV	id	ka	th	fa	vi	pl	cs	de	no	sl	sv	nl
id-Indonesian	5												
ka-Georgian	15	←											
th-Thai	19	←	-										
fa-Persian	23	←	-	-									
vi-Vietnamese	30	←	-	-	-								
pl-Polish	47	←	-	-	-	-							
cs-Czech	70	←	-	-	-	-	↑						
de-German	79	←	←	←	←	←	-	←					
no-Norwegian	81	←	-	-	-	-	-	-	↑				
sl-Slovenian	81	-	-	-	-	-	-	-	↑	-			
sv-Swedish	87	←	-	-	-	-	-	-	-	-	-		
nl- Dutch	100	←	←	-	-	-		←	-	-	←	-	

The Mann-Whitney U test confirmed H3, the idea that the distribution of high-quality content creator activity in namespaces was markedly different for each language version in view of the IDV value. In addition, the characteristics of contributors according to categorised forms of activity selected the most and least active among the users of the surveyed language versions of Wikipedia. In the case of the Indonesian version, only once did users gain an advantage, in the technical space (the domain of collectivist cultures) over the German version. The Polish version performed less well against 9 out of 11 language versions, only gaining an advantage over Indonesian and Czech in workspaces. The undeniable favourite in this list is the version of the German Wikipedia, which has only been off the podium podium of the greatest activity in technical spaces. The German version has an overwhelming advantage in the number of active users compared to the other language versions, which suggests that the number of users is a strong indicator. However, in the present study, the characteristics of individual user statistics were used. The overall size of a given version was much less significant. The number of users whose activity was recorded was determined by the frequency of co-authorship in the articles selected.

Therefore, the number of respondents for the 12 language versions varied for each of them, ranging from 23 to 41 users.

5. Discussion and Limitations

It seemed reasonable to assume that high quality was produced primarily by a small group of experts. It did not work for the form of work characteristic of creating the content of Wikipedia articles. The higher number of featured article editors in individualistic cultures may obviously be due to the much higher number of active users overall. However, it should be taken into account that editors of an equal number of articles were compared. Perhaps the reason for such a large difference in the number of editors lies in the greater need to assert one's own individual position in the Wikipedian community, which is attributed to individualistic cultures.

A low or high IDV does not give unambiguous results in relation to the share of featured articles in individual language versions.

The substantive space showed the greatest polarisation of the language versions of Wikipedia in terms of the cultural dimension of individualism vs. collectivism. Activity in substantive spaces is the domain of individualistic cultures. Collections of language versions with high and low IDV in technical spaces, unlike those studied in the case of substantive spaces, are the domain of collectivist cultures. Content promotion spaces remain dominated by individualistic cultures. The results of hypotheses verification are presented in table 7.

Table 7: Results of hypotheses verification

H1 – According to the preferred models of work, the featured articles of language versions with high IDV are created by a small group of people, as opposed to cultures with low IDV.	not confirmed
H2 – The index of individualism in a given culture translates into a higher or lower share of featured articles in particular language versions.	not confirmed
H3 – The distribution of activity of high-quality content creators in namespaces is clearly different for each language version, in view of the IDV value.	confirmed

A characteristic feature of all featured articles, regardless of language version, is the disproportionately large share of individual people in their creation. This study confirms the earlier results of Feldstein (2011), among others. Feldstein pointed out that at the article level, creating a text is more like the traditional creative process of an individual author. Previously, Kittur and Kraut (2008) showed that the quality of an article was positively influenced by the cooperation of a small group of editors, who create an article which is subsequently joined by a larger group. In the present study, the contribution of an outstanding editor for 10 language versions averaged 70-80% of the text. For Indonesian and Norwegian, the figure was over 80%. In the Polish version, on average, 90% of the contribution to the content of the article was made by one user. Therefore, the bottleneck in creating high-quality content in the different language versions of Wikipedia is the involvement of experts. The number of featured articles on a given topic can be a clear indicator of the presence of experts among Wikipedia users.

In the analysis of user activity in individual namespaces, there are individual exceptions to the rule of "superactive" people. Out of nearly 400 cases, 3 "superactive" users were found. If a given user has too many edits (maximum 650,000), the counter is simplified, and there is no preview of activity in individual namespaces.

Creating Wikipedia articles can take a long time. The ability to make changes to articles makes it an ongoing process, although many entries remain unchanged for years. In the case of featured articles, we are dealing with content that can be considered highly comprehensive. The primary difficulty that applies to all people exploring Wikipedia statistics is the collection of data at any given point in time. The data is becoming less relevant day by day. If these are small changes, e.g. in the number of editors of a given article or their contributions, the differences will not be significant. However, when more general data is updated every few months, and in some cases once a year or even at longer intervals, it can have an adverse effect on the reality of the analysis. It is therefore advisable to check the validity of the overall data not only at the beginning but also at the end of the study.

Language versions of Wikipedia sometimes use namespaces in individual ways, for activities that are assigned to other spaces. In addition, each language version can define its own namespaces with no equivalents in other language versions. Of the 12 versions of Wikipedia studied, the Indonesian version is example of this. Indonesian has a Story in its namespaces, which contains catalogues of photos from the phone. Such differences in namespaces require a qualitative study.

6. Conclusions

Statistically, the study showed significant differences in the form of author and editor engagement in creating high-quality articles. One should be cautious about generalising conclusions to cover other versions of Wikipedia. However, the results obtained do provide grounds for concluding that the quality of content depends on Hofstede's cultural indicators. In future research, one may assume that representatives of individualistic cultures are superior to representatives of collectivist cultures in the creating high-quality knowledge.

This paper is an extension of the research on the community involved in the sharing of knowledge on the Internet. The determinants of user activity, resulting in the creation of high-quality content, are presented. It would be advisable to include a qualitative analysis in the quantitative study presented, to highlight the differences in the use of namespaces by individual language versions. Qualitative data would give a more precise definition of the cultural space of representatives of different language versions. This could result in the emergence of a new cultural dimension that would more accurately describe the social reality on the Internet.

References

Alonso, E. and Robinson, B.J. (2016) "Exploring Translators' Expectations of Wikipedia: A Qualitative Review", *Procedia - Social and Behavioral Sciences*, No. 231, pp 114-121.

Feldstein, A. (2011) "Deconstructing Wikipedia: Collaborative Content Creation in an Open Process Platform", *Procedia - Social and Behavioral Sciences*, No. 26, pp 76-84.

Giles, J. (2005) "Internet encyclopaedias go head to head", *Nature*, Vol 438, No. 15, pp 900-901.

Hofstede, G., Hofstede, G.J. and Minkov, M. (2010) *Cultures and Organizations. Software of the Mind. Intercultural Cooperation and Its Importance for Survival*, McGraw-Hill Professional.

https://en.wikipedia.org/wiki/List%5fof%5fWikipedias

https://en.wikipedia.org/wiki/Wikipedia:Compare_criteria_Good_v._Featured_article

https://meta.wikimedia.org/wiki/Wikipedia_featured_articles

Iba, T., Nemoto, K., Peters, B. and Gloor, P.A. (2010) "Analyzing the Creative Editing Behavior of Wikipedia Editors Through Dynamic Social Network Analysis", *Procedia Social and Behavioral Sciences*, No. 2, pp 6441-6456.

Jemielniak, D. (2019) "Wikipedia: Why Is the Common Knowledge Resource Still Neglected by Academics?", *GigaScience*, Vol 8, No. 12, giz139.

Kane, G.C. (2011) "A Multimethod Study of Information Quality in Wiki Collaboration", *ACM Transactions on Management Information Systems* (TMIS), Vol 2, No. 1, pp 1-16.

Kittur, A. and Kraut R.E. (2008) "Harnessing the wisdom of crowds in wikipedia: Quality through coordination", ACM Conference on Computer Supported Cooperative Work, (San Diego), pp 37-46.

Kittur, A., Pendleton, B., and Kraut, R.E. (2009) "Herding the cats: the influence of groups in coordinating peer production", Proceedings of the 5th international Symposium on Wikis and Open Collaboration, pp 1-9.

Konieczny, P. (2021) "The Case for Institutional Support: It's High Time for Governments and University Administration to Actively Support Wikipedia", *The Journal of Design, Economics, and Innovation*, Vol 7, No. 2, pp 187-196.

Kukowska, K. (2022) "Knowledge Sharing Co-operative Structures – The Case of WikiProjects", Proceedings of the 23 nd European Conference on Knowledge Management, Naples, Italy, pp 696-704.

Kukowska, K. (2023) "Collaborative Structures in WikiProjects. Wikipedia Users' Activity Case Study", Matos, F., Rosa A. (eds.), Proceedings of the 24th European Conference on Knowledge Management, Vol 24 No. 1, pp 760-768.

Kukowska, K. and Skolik, S. (2021) "Wikipedia as a Space for Collective and Individualistic Knowledge Sharing", Proceedings of the 22nd European Conference on Knowledge Management, Coventry, United Kingdom, pp 459-466.

Lewoniewski, W., Węcel, K. and Abramowicz, W. (2016) "Quality and Importance of Wikipedia Articles in Different Languages", Proceedings: Information and Software Technologies: 22nd International Conference, Druskininkai, Lithuania, pp 613-624.

Lewoniewski, W., Węcel, K. and Abramowicz, W. (2019) "Multilingual Ranking of Wikipedia Articles with Quality and Popularity Assessment in Different Topics", *Computers*, 8, 60, pp 1-31

Lewoniewski, W., Węcel, K. and Abramowicz, W. (2023) "Understanding the Use of Scientific References in Multilingual Wikipedia across Various Topics", *Procedia Computer Science*, 225, pp 3977-3986.

Mesgari, M., Okoli, Ch., Mehdi, M., Nielsen, F.Å. and Lanamäki, A. (2015) "'The Sum of All Human Knowledge': A Systematic Review of Scholarly Research on the Content of Wikipedia", *Journal of the Association for Information Science and Technology*, Vol 66, No. 2, pp 219-245.

Nemoto, K. and Gloor, P.A. (2011) "Analyzing Cultural Differences in Collaborative Innovation Networks by Analyzing Editing Behavior in Different-Language Wikipedias", *Procedia - Social and Behavioral Sciences*, 26, pp 180-190.

Pfeil, U., Zaphiris, P. and Ang, C. (2006) "Cultural Differences in Collaborative Authoring of Wikipedia", *Journal of Computer-Mediated Communication*, 12, pp 88-113.

Ren, Y., Chen J. and Riedl J. (2016) "The impact and evolution of group diversity in online open collaboration", *Management Science*, Vol 62, No. 6, pp 1668-1686.

Stvilia, B., Al-Faraj A., and Yi, Y. J. (2009) "Issues of cross-contextual information quality evaluation – The case of Arabic, English, and Korean Wikipedias", *Library & Information Science Research*, Vol 31, No. 4, pp 232-239.

Szajt, M. (2014) *Przestrzeń w badaniach ekonomicznych*, SWWZPCz, Częstochowa.

Zhang, H., Ren, Y. and Kraut R.E. (2020) "Mining and Predicting Temporal Patterns in the Quality Evolution of Wikipedia Articles", Proceedings of the 53rd Hawaii International Conference on System Sciences, pp 3971- 3980.