

Article

# Cryptocurrencies Perception Using Wikipedia and Google Trends <sup>†</sup>

Piotr Stolarski \* , Włodzimierz Lewoniewski and Witold Abramowicz

Department of Information Systems, Poznań University of Economics and Business, 61-875 Poznań, Poland; Wlodzimierz.Lewoniewski@ue.poznan.pl (W.L.); Witold.Abramowicz@ue.poznan.pl (W.A.)

\* Correspondence: Piotr.Stolarski@ue.poznan.pl

† This paper is an extended version of our paper published in Business Information Systems Workshops (BIS 2019), 26–28 June 2019, Seville, Spain.

Received: 26 March 2020; Accepted: 22 April 2020; Published: 24 April 2020



**Abstract:** In this research we presented different approaches to investigate the possible relationships between the largest crowd-based knowledge source and the market potential of particular cryptocurrencies. Identification of such relations is crucial because their existence may be used to create a broad spectrum of analyses and reports about cryptocurrency projects and to obtain a comprehensive outlook of the blockchain domain. The activities on the blockchain reach different levels of anonymity which renders them hard objects of studies. In particular, the standard tools used to characterize social trends and variables that describe cryptocurrencies' situations are unsuitable to be used in the environment that extensively employs cryptographic techniques to hide real users. The employment of Wikipedia to trace crypto assets value need examination because the portal allows gathering of different opinions—content of the articles is edited by a group of people. Consequently, the information can be more attractive and useful for the readers than in case of non-collaborative sources of information. Wikipedia Articles often appears in the premium position of such search engines as Google, Bing, Yahoo and others. One may expect different demand on information about particular cryptocurrency depending on the different events (e.g., sharp fluctuations of price). Wikipedia offers only information about cryptocurrencies that are important from the point of view of language community of the users in Wikipedia. This “filter” helps to better identify those cryptocurrencies that have a significant influence on the regional markets. The models encompass linkages between different variables and properties. In one model cryptocurrency projects are ranked with the means of articles sentiment and quality. In another model, Wikipedia visits are linked to cryptocurrencies' popularity. Additionally, the interactions between information demand in different Wikipedia language versions are elaborated. They are used to assess the geographical esteem of certain crypto coins. The information about the legal status of cryptocurrency technologies in different states that are offered by Wikipedia is used in another proposed model. It allows assessment of the adoption of cryptocurrencies in a given legislature. Finally, a model is developed that joins Wikipedia articles editions and deletions with the social sentiment towards particular cryptocurrency projects. The mentioned analytical purposes that permit assessment of the popularity of blockchain technologies in different local communities are not the only results of the paper. The models can show which country has the biggest demand on particular cryptocurrencies, such as Bitcoin, Ethereum, Ripple, Bitcoin Cash, Monero, Litecoin, Dogecoin and others.

**Keywords:** cryptocurrencies; wikipedia; bitcoin; information demand; information quality

## 1. Introduction

The purpose of the research is to explore how diversified decentralized cash systems are presented and characterized in the largest open-source knowledge base. During this study, a number of research questions (RQs) were raised. They are listed below:

1. Are the articles that describe cryptocurrencies within Wikipedia emotionally well balanced? To what extent are they neutral in their claims? (RQ1)
2. Is there any association between the sentiment of Wikipedia articles about crypto coins and their overall quality? (RQ2)
3. Whether popular search engine statistics show similar patterns of interest as the visits in Wikipedia about cryptocurrencies? (RQ3)
4. How can one model the popularity of particular cryptocurrencies based on demand for information on the Internet? Is it possible to track this popularity on a geographical basis? (RQ4)
5. Is it possible to bind the national attitudes towards the crypto economy with the popularity of cryptocurrencies in particular countries? How different in particular countries is the legal approach towards cryptocurrency technology? (RQ5)
6. Why only several cryptocurrencies are described in the world encyclopedia? How variable is the crypto economy subject matter presented in Wikipedia? (RQ6)

Nowadays huge amount of content in Internet created by individuals helps to provide research in different fields: medicine [1], tourism [2], marketing [3] and others. There are various possibilities for the Internet users to create such content. One of the most popular examples of such services is Wikipedia. This free collaborative knowledge base contains social and behavioral data, which already proven to be useful for socio-economic forecasting [4]. For example, data from Wikipedia can be used for predicting movie box office success [5] or moves on stock market [6].

The elaborated topic is important because both Wikipedia and cryptocurrency technology are now used on a mass scale and internationally. There is also a strong need for a trustful and independent information about particular cryptocurrencies. Wikipedia has been perceived for a long time as a knowledge source of lesser reliability despite its popularity. This negative attitude and unenthusiastic sentiment concerning its trustworthiness have evolved in time towards more positive opinions. One has to bear in mind that Wikipedia is the most important crowdsourcing web portals on the modern Internet. Therefore, it should be treated as the major mean for modeling Internet users' behaviors. Especially that Wikipedia provides access not only to content, but also to metadata about history of editing, readers, actions of the editors and other potentially useful data.

Wikipedia articles often appears on the top positions in search results in Google, which provide special tool for popularity analysis of search queries—Google Trends [7]. This tool was used as additional source of data, to analyze demand for information about cryptocurrencies in different countries and for different time periods.

The first cryptocurrency started at the beginning of 2009. It was designed to function in analogy to normal (fiat) money. However, it had some exceptional features, being completely digital, third-parties independent and anonymous to a certain degree. After a decade, together with other, similar projects, it is a basis of an international economic system. The size of this system is also comparable to medium-sized national economies. The number of cryptocurrency projects that evolved during this time is beyond two thousand. Nevertheless, only a small amount of them is of high popularity.

In this research, five relationships between information and metadata contained in Wikipedia and the particular instances of cryptocurrencies were studied. The described analytical instruments allow formulating assertions about the state of cryptocurrency technologies and the position of specific cryptocurrency projects in the realm of these technologies. Specifically, the study involves:

- The quality of crypto coins and their descriptions in Wikipedia.

- The extension of the formulation of the cryptocurrencies' popularity model that allows the estimation of potential users in particular geographic locations.
- The construction of a model that confronts the crypto coins' popularity with the legality of their use in certain jurisdictions.
- The analysis of Wikipedia articles' dynamics and its comparison with numerous cryptocurrency features.

The contribution of this study consist of conducting cryptocurrency-related Wikipedia articles deletion analysis, preparing cryptocurrency-related Wikipedia articles sentiment ranking, extending the cryptocurrency popularity model proposed in [8] and providing the model that confronts the national cryptocurrency popularity data with their legality.

The methodological approach taken in this research uses the design science research (DSR). The DSR is a framework that allows to systematize the smooth transition from theoretical background to materialized empirical artifacts. It is especially well-suited to the IT-related studies. Additionally, some statistical methods are supplementary employed.

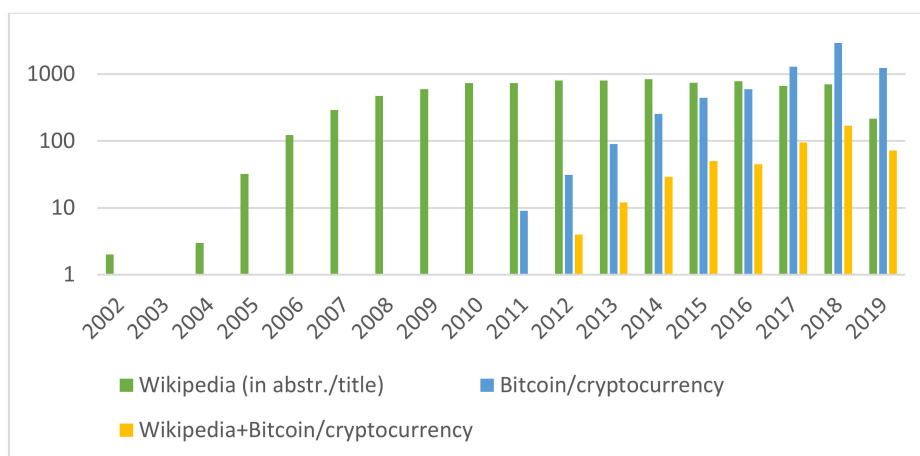
Wikipedia should be deemed as an unbiased source of knowledge and information for cryptocurrency technologies. The idea behind every encyclopedia is to present facts about objects of diversified types in the form of articles. These articles reflect community knowledge from any domain. When comparing Wikipedia to any other encyclopedic effort, it is exclusive for several reasons. It is the largest, international, free, open and collaboratively written web-based knowledge source. It has more than 300 language versions which in total exceeds 52 million articles. There is also an informative category describing cryptocurrencies and related concepts.

The details about Wikipedia treated as the source of information are covered in Sections 3.1 and 4.1.1 of the article. In the Section 2, a reference is given to other researches that may be considered relevant to the presented one. Consequently, Section 3.2 gives an in-depth description of the technical side of data extraction from Wikipedia. All the presentations of the mentioned analytical instruments are organized in Section 4 of the paper. The implications of received results are discussed in Section 5.

## 2. Literature Review

There exist a small but constantly growing number of papers that involve both subjects of cryptocurrencies and Wikipedia. Some authors of the mentioned articles also turn their attention to other Internet knowledge and data sources. However, most of the earlier conducted researches focus on the associations between market time-series and the particular indicators based on crowd media data (e.g., [9,10]). In opposition to these texts this paper focus rather on other aspects of crypto coins than strict market data. A similar approach was taken in a recent study presented in [11].

Figure 1 presents the distribution of articles related to the topics of interest of this paper. All the considered texts have been published in the following years since 2002—that is the creation of Wikipedia. This is a demonstrative graph that is aimed at giving orientation in numerical relations between scholarly contributions within the topics of Wikipedia and cryptocurrencies. The data come from the Scopus database. The articles devoted to Bitcoin started to appear in 2009. The third dataset represents the papers combining both the topic of Wikipedia and cryptocurrencies. The specifics of Wikipedia makes it a little problematic to set a precise number of involved scientific texts. The reason is that it may be included in the content of scientific papers in two roles: an object of research or as an information (reference) source. As the latter case is prevalent the search results were narrowed to the title or abstract of the papers.



**Figure 1.** Number of publications in subsequent years (logarithmic scale).

Researches use Wikipedia regularly not only as a source of information but also as the main object of study. The researches on Wikipedia are especially focused on the topics of information quality and reliability of Internet services as an information source. The paper of Rowley and Johnson [12] is an example of a study that aims to recognize the important elements that are the basis of the trustworthiness assessment process of an Internet information source. This study consist of two phases and Wikipedia was chosen as the object of examination. The same study also offers insights into the process of how these elements are used to evaluate the web source. Most the factors that were found are consistent with previous reports on the topic.

The authors of [13] present two main findings. First, they empirically show that Wikipedia is a very popular knowledge source for students. Second, the authors highlight that the use of the knowledge form the on-line free encyclopedia is done without any particular consideration of the obtained information quality. It is an improbable situation that the students further verify the knowledge from Wikipedia. It implies that this Internet source is treated often by the young generation as an ultimate trusted provider of knowledge although it is quite a subjective point of their view.

The trustworthiness of Wikipedia is also admitted by another study [14]. This research is more precise in its assessment of the popularity of this source of knowledge. The authors indicate that a third of the tested college group extracted facts from this source of information. Although the paper does not deal with the issue of information verification, another text [15] suggests that part of the students show the behavior of critical reading of the online encyclopedia content. This observation is in contrast to the one mentioned earlier [13]. Another conclusion of [15] is that when the trustworthiness of an encyclopedic article was dubious the study subjects used different quality measures to further determine the reliability of the content.

Data from Wikipedia can be used not only to analyze demand for information on specific topics, but also can help to predict success of products and moves on stock markets. Research of Mestyán et al. [5] presented a predictive model of financial success of movies based on such measures as number of authors and pageviews of Wikipedia articles. Scientific work by Moat et al. [6] showed how Wikipedia data of user activity and popularity of articles can improve the decision-making process in stock market. There is also a study based on Wikipedia usage that created a model for tourism demand forecasting [16].

Other source of data that was used in this study is Google Trends. It allows to predict sales of the different products: cars [17], telecom [18], fashion [19] and others. However, Google Trends gives only relative values of the popularity of search queries in the scale from 0 to 100. Additionally, this tool has limitation on getting data for long periods of time and sometime problems of query unification for different languages (when there were various spelling variants for the same topic). This last problem can be solved in Wikipedia using semantic connections between the language versions.

The article of Kristoufek [9] is an example of one of the very few researches that made the comparison between Bitcoin market time-series Google Trends data and some Wikipedia statistics. The selection of the data sources is not coincidental. Few later studies that are elaborated below process data from a similar or expanded set of sources. The experience from the research is that a correlation may be found between Bitcoin capitalization and Google queries. The importance of this report lays in its innovativeness by presenting a new front of inquiries seeking relationships linking two significant Internet trends: Social Web and cryptocurrencies.

The relations between Bitcoin and various other time-spun sets of data are also studied in [20]. In this research, the behavior of the cryptocurrency market for Bitcoin is compared with several traditional real economy indicators as well as digital economy news feeds. The collection of processed datasets encompasses fiat currencies exchange rates, stock exchange indices, social media news, as well as queries on Wikipedia, Google search engine and Twitter posts. Numerous econometric techniques have been employed during the research. The authors found the existence of a long-run negative bound between S&P500 and the Bitcoin value. The relation shows the substitution between stocks and Bitcoin as an alternative investment instrument. The investors accumulate their capital according to the prospect of the global state of the economy. Once the outlook is pessimistic the alternatives grow on importance.

The approach presented in the mentioned articles ([9,20]) is also taken in the research of [21]. The last paper is however much better supported by the economic theory. Beside a large number of determinants that potentially may influence both fiat currencies and crypto assets, the article fits the explanation of the results of obtained results into the well-established Barro's model for gold. The fundamental motivation for the research is to discover the forces behind the pricing mechanism. One of the specific factors that according to authors should capture the investment attractiveness of the examined cryptocurrency is the daily number of views of the article about Bitcoin on Wikipedia. The authors, however, note that such a measure has a flaw which is a failure to differentiate on what is the exact purpose of the demand for the information. The motives that drive the behavior of actors visiting the particular article on the online encyclopedia may be manifold and range from investors to technology (potential) users. According to the authors the long run macro-financial developments have no measurable impact on Bitcoin prices. Nevertheless, certain factors such as the demand-supply balance and Bitcoin attractiveness that have a considerable influence on the cryptocurrency market value. Another conclusion is that the correlation strength among the influential elements is not constant in time.

Kristoufek extended his initial research of 2013 into another article. Although the aim and motivation for the following study [22] are different, the data used are very similar to that from the previous examination [9]. The main motivation for the paper dated 2015 is the assessment of the degree to which Bitcoin is a purely speculative asset and what are the most important factors that impact the volatility of Bitcoin price. The author categorizes these factors into technical, fundamental and speculative ones. The text analyzes comprehensively the possible ways in which the elements may impact cryptocurrency market prices. Not only are temporal changes taken into account but also the partition on short- and long-term stimuli is introduced. The latter is based on signal frequency distribution by utilizing the continuous wavelets framework. Search engines are one of the groups of data series that impact on Bitcoin market is measured. The group includes two data sets: weekly Google Trends and daily Wikipedia visits, both accumulated to represent the term "Bitcoin".

Social media were used for algorithmic trading on Bitcoin [23] and altcoins price prediction [24]. Some authors used microblogs sentiment analysis related to cryptocurrencies [25–27].

Two main facts may be reconstructed after reading up the quoted literature. First, a number of researches explore the issues of interconnections between the cryptocurrency market and crowd media together with Google Trends. Wikipedia is among these prominent crowd media. This means that their authors are aware of the potential of the relationships to study. Nevertheless, the studies conducted up to this point merely scratched the broad capabilities of examining possibly fruitful aspects that



this source represents. At least two aspects are omitted by most studies. That is the broadness of the cryptocurrency market with thousands of coins and tokens. Another aspect is the internationalization is a crucial attribute of the free online encyclopedia. As of 2019, there exist beyond 300 language editions which are independently managed, and which offer independent content.

Second, there is an ongoing discussion on the quality of articles and the credibility of Wikipedia as a source of knowledge. One thing is however beyond doubt, which is the shift of critical attitude towards this issue over time. The encyclopedia community has made definitive progress on the way to improve the editorial standards so that Wikipedia is now much more believable than it used to be in the past. Unlike most similar works in this area, data from various language versions of Wikipedia were used. Moreover, a wider set of measures for cryptocurrencies analyses was extracted.

### 3. Method

The research is based on a mixed methodology. It uses the design science method as presented by [28]. Additionally, some statistical methods are employed as well. The main method used is widely accepted and frequently employed in the research related to information systems and information technologies. [29] is built on evaluation of 26 journal articles that are conducted according to the method. Another paper [30] conduct a bibliometric content analysis of 362 sample texts that apply design science research (DSR) approach. These papers come from 14 journals.

The DSR framework is the backbone of the study. It is used to rigorously sort the subsequent research tasks. It also introduces the methodological soundness by breaking down the operationalization into three cycles. The rigor cycle has a foundational role. The systematic and critical literature review is the key point to specify the current knowledge base and to develop proper research questions. The following is the design cycle. At this stage, the general research questions undergo the process of a careful breakdown into hypotheses. Additionally, the most suitable modeling methods are selected. The datasets are prepared and cleaned. The next stage is the relevance cycle. In this cycle, the models are examined, and the research questions are answered. Because of the cyclic nature of the DSR method, some of the steps may be recurrent to expand the already obtained results and to finetune the models. The identification of the environment units that are susceptible to the effects of the examination is also part of the relevance cycle. The models and the datasets obtained during the study constitute the artifacts of the design cycle.

Quantitative methods are used to build models. At the same time, they measure the level of relations between particular objects of interest. Finally, they allow formulation and prove some of the hypotheses.

Parts of the presented study and the article can be directly mapped onto the framework. In the introduction the general research questions were formulated. The literature review section is to initially populate the knowledge base. These parts constitute the rigor cycle. In the Sections 3.1 and 3.2 there is a description of the process of gathering the data to obtain the datasets. These datasets are used in order to prepare rankings and models that are described in details in the results section. All these activities are covered by the design cycle of the framework. Given the iterative nature of this cycle, the results obtained from the cryptocurrency popularity model are used further in the popularity-legality model. The items that make up the final—relevance cycle are presented in the Implications section and partially in the Results section. These are the answers to the research questions and results discussion as well as verification of the models (Sections 4.2.3 and 4.3.1).

#### 3.1. Information Extraction from Wikipedia

Wikipedia is one of the most popular collaborative knowledge bases on the Internet—it is placed on 5th place in the global ranking of the most visited websites. Information in this free encyclopedia is created and edited by users from different countries in over 300 language versions. Wikipedia contains over 52 million articles about different subjects. In each language chapter of this encyclopedia users can create and edit articles about specific topics independently. Therefore, some important subjects for

the local community are placed in a limited number of language versions and even the most developed English Wikipedia with over 6 million articles does not contain all information about some locally relevant personae or places that do have entries in language-specific versions. Additionally, if the same topic has entries in more than one language, it does not mean that other language versions must have analogical information. That is why there are often differences in information quality between language versions that describe the same subject.

In this research, Wikipedia is used as a source for information extraction on three separate levels of abstraction. At the first level, there exists a content layer [31,32]. It is represented by categorized articles divided into structured units of text. Each article and its subsequent sections are focused on a specific topic. The second level consists of metadata and statistics accompanying every article and web page on the portal. Such data as the number of views, editorial history, traces of users' interaction, comments and discussions are available within this group. The third level of abstraction is made up of particular indicators or measures that are based on the former levels of available data [33,34].

The articles in Wikipedia usually are aligned to one or more categories. The categories have a graph-like structure. It means that a category may be further divided into subcategories. The articles themselves are a sort of structured documents where the text is divided into entitled sections and subsections. Moreover, the articles use formally described templates and may use some additional predefined patterns. Specific parts of the articles are annotated with quotes and bibliographic information which establish a better quality of the provided information as well as improve its reliability. An article may also contain fully structured information in the form of infobox that resembles the object-oriented or relational database facts presentation.

Wikipedia is a fully open and knowledge system. The Wikimedia Foundation Inc. is formally responsible for maintaining the encyclopedia project as well as all connected other initiatives. The foundation publishes all the information contained on the Wikipedia portal as well as all related metadata in a timely fashion. The snapshots of the encyclopedia are available in several dump files [35]. The analysis of these files allows accessing all the knowledge and making analysis of metadata as well as a calculation of measures or indicators based on the provided information [33].

### 3.2. Process of Metadata Transformation

A differentiated set of methods can be used to find Wikipedia articles related to specific cryptocurrency. Those methods are based on:

- articles content and its metadata (for example Wikipedia category, name of the Infobox),
- related item in DBpedia and Wikidata.

There are different possibilities and techniques to get measures values of the Wikipedia articles. The vast majority of the measures can be extracted from the Wikipedia database dumps [35]. A list of important files from the latest dump of English Wikipedia is presented in Table 1 with a brief description of parameters that can be extracted from their content [34]. As seen from the list the mentioned dump files are an abundant source of vastly diversified measures and give the opportunity to straightforwardly extract their values. Nevertheless, not all of the interesting measures can be extracted from these files. For example, to obtain the number of webpage visitors for each article, it is necessary to send a request to Wikipedia API [35]. Measures related to popularity can be extracted from other databases or API services.

**Table 1.** Parameters of Wikipedia persisted in database dumps.

Source	Content
enwiki-latest-pages-meta-current.xml.bz2	Recombines all pages (including articles), current versions only. This file is used for obtaining a majority of the article measures
enwiki-latest-pages-articles.xml.bz2	Consists of articles, templates, media/file descriptions and primary meta-pages. It can be used also for obtaining a majority of the articles' measures (excluding statistics from discussion pages).
enwiki-latest-pagelinks.sql.gz	Wiki page-to-page link records. Used for network measures—for example, incoming links from other articles.
enwiki-latest-categorylinks.sql.gz	Wiki category membership link records. It can be used for the category count measure.
enwiki-latest-externallinks.sql.gz	Wiki external URL link records. It can be used to measure external link count.
enwiki-latest-stub-meta-history.xml.gz	Contains only historical revision metadata. It can be used to extract the number of editors from different groups (bots, anonymous users, administrators, etc.) and also the number of edits of various types (e.g., minor edits, edits comments).
enwiki-latest-iwlinks.sql.gz	Interwiki link tracking records. It can be used to extract the number of unique internal links (links to other Wikipedia articles).
enwiki-latest-templatelinks.sql.gz	Wiki template inclusion link records. Used for templates count measure, also it is possible to check if an article has an infobox
enwiki-latest-page.sql.gz	Base per-page data (id, title, old restrictions, etc.). It can be used to extract last edit time, page length in bytes.
enwiki-latest-imagelinks.sql.gz	Wiki media/files usage records. It can be used to measure image count.

From the point of view of a thorough study, complete research material is an important aspect. The domain of cryptocurrencies is represented in Wikipedia both by single articles that deal with the description of particular decentralized money systems and specific topics of broader nature. In the research, a comprehensive list of articles was obtained from two sources that include Wikipedia categorization and the information from DBpedia and Wikidata [34,36,37]. The latter ones are examples of semantic knowledge bases related to Wikipedia and its content. Another aspect is the quality of the research material.

The WikiRank [38] measure is a good example of processing and using the data that can be obtained from Wikipedia dump files. The WikiRank project use some of the important measures to assess the quality of the Wikipedia articles using a synthetic measure. Different parameters can be related to such quality dimensions as completeness, credibility, objectivity, readability, relevance, style and timeliness [34].

The quality in Wikipedia is a topic broadly described in scientific works [33]. Some of the researches focus on methods for automatic predicting of the article's quality of Wikipedia. Each study usually used its own set of measures and specific algorithm to build a model to solve this task. For example, text length, number of the references, images, sections, popularity are among the most important parameters. In Section 4.1.4 the WikiRank measure will be used in order to compare it with the results of one of the conducted analyses.

The international nature of Wikipedia allows preparing a language-related comparison of the articles. One may count the number of language versions of a given article that is related to certain cryptocurrency. It is obvious that the more important entry the more language versions it will have. Of course, some local phenomena may be quite important for a local (national) community which should result in an entry that has little counterparts in other language versions. Most crypto coins are part of the global crypto-economy and as the electronic markets are highly globalized there is a little



place—at least so far—for national or localized cryptocurrencies or tokens. Petro is an alter-example of this. Table 2 presents the crypto coins in which articles in Wikipedia have the largest number of language versions and thus are most internationalized. Additionally, some specific cases are indicated. These are mainly the examples of encyclopedic entries that were deleted. There will be a reference to these deleted instances in Section 4.1.2.

**Table 2.** Ranking of cryptocurrency articles with the largest number of language versions.

Name	Number of Languages	Capitalization (US \$)
Bitcoin	99	\$138,257,959,928
Litecoin	44	\$3,087,535,172
Ethereum	44	\$21,028,815,007
Dogecoin	32	\$229,265,532
Monero	27	\$1,748,932,689
Ripple	24	\$18,450,540,329
Libra	24	n/d
BitcoinCash	21	\$7,679,814,861
IOTA	19 (D)	\$1,372,123,792
Dash	18	\$652,067,615
Onecoin	17	n/d
Zcash	12	\$312,369,154
Namecoin	10	\$4,948,692
Petro	10	n/d
Cardano	D	\$789,252,402
Tron	D	\$771,475,812
TOTAL	401	

At the moment when the text is written, there are currently 45 articles describing individual cryptocurrencies on the English version (EN) of Wikipedia. It is not much taking into account the global number of cryptocurrencies and because English Wikipedia is the most influential and largest language version of the open encyclopedia. The described cryptocurrencies make less than two per thousand of the total crypto coins in the market. However, if all language versions of Wikipedia will be included, 77 cryptocurrencies are described in at least one of them.

On the other hand, it has to be taken into account that Wikipedia has worked out numerous standards and quality assurance procedures which result in eliminating or at least discouraging descriptions of entries that are questionable, not factual or possess other characteristics that make such an entry disputable. One particular factor that is vital for the excellence of the Wikipedia articles is the write-lock mechanism. The rules of the mechanism are different for every language version. Nevertheless, the main reason and fundamental aspects are unchanged. Although Wikipedia is an open encyclopedia that potentially attracts both readers and editors for restriction-free sharing of knowledge, in the case of part of articles the access for editing actions is not completely open. The English Wikipedia governs itself with a ten-fold level of action permissions.

## 4. Results

### 4.1. Wikipedia Cryptocurrency-Related Articles Sentiment and Deletion Analysis

#### 4.1.1. Wikipedia Editorial Standards

The main approach of Wikipedia is that it is open to be edited by any volunteering party. At the same time, all the contents are expected to meet some quality expectations which are reflected in rules. For example, Wikipedia articles must be written with the appropriate style and present facts from a neutral point of view [33].

The creation and work on articles are based on the assumption that the editors will strive to achieve a stable version by achieving consensus. The more technical side of Wikipedia to enforce

certain behaviors or omissions uses an extensive article security system. The system is independent in each language version. The same applies to policies on encouraging improvement in the quality of content.

#### 4.1.2. Articles Dynamics

The changes in the number of cryptocurrency-related articles in particular language editions of Wikipedia between years 2020 and 2018 are provided in Table 11. Whereas, in Table 2 a sample of most valuable crypto coins was presented together with their reference values of Wikipedia popularity indicators. Some of the cryptocurrencies are listed with the “D” symbol, which stands for “deleted”. In practice, there are at least three possibilities for a Wikipedia article page to be deleted. The most typical scenario is that the article is removed after nomination and proper discussion. It means that there is a formalized process that is designed in order to filter out unwanted or troublesome “encyclopedia” entries. The second scenario is intended for alert situations. It is named by Wikipedia users as “speedy deletion”. Fast deletion is a procedure that involves minimum formality and has abrupt consequences. In this case, the removal of the entry is done without any discussion conducted. The third of the mentioned possibilities is a deletion of a page, which itself was a redirection to other removed articles.

An important issue that should be noted is that every language version of Wikipedia has its autonomy in designing and providing means of improving content trust and quality. It implies that each version has its own procedures and sets of rules that would affect filtering and removing of poor or needless articles. Additionally, it also signifies that the decision that given entry is not worthy of being kept in one language version does not necessarily lead to similar rejective actions in other versions. Hence, it is not an axiom that the most popular, i.e., the one with the largest number of entries language version will always have the fullest range of articles (topics).

In the study, the cases of removal decisions in English Wikipedia were investigated. In total, over 300 pages (including articles, redirections) about cryptocurrencies were deleted. One hundred and fifty-five entries—about 52%—were removed using speedy deletion procedure [39]. Table 3 shows the indicated reasons for fast-deletion cases, accompanied by a number of items that correspond to specific cryptocurrencies.

**Table 3.** Speedy deletions of crypto coin articles with the criterion of fast deletion appliance.

Symbol	Criterion	Num. of Articles
G11	Unambiguous advertising or promotion	48
A7	No indication of importance (people, animals, organizations, web content, events)	21
G8	Pages dependent on a non-existent or deleted page	18
G12	Unambiguous copyright infringement	16
G6	Technical deletions	14
G7	Author requests deletion	7
G3	Pure vandalism and blatant hoaxes	6
A3	No content	5
G4	Recreation of a page that was deleted per a deletion discussion	4
R2, R3, G5, A11, A1, U5, A10, G2	Other reasons	16

Another 93 articles were deleted after nomination and discussion [40]. Table 4 presents the selected article that was deleted as a result of such discussion with information about number of visits of the discussion page, number of authors that participated in discussion and number of nominations for deletion.

**Table 4.** Number of authors and deletion nominations for particular articles.

Name	Visits	Authors	Nominations
IOTA (technology)	6092	16	1
Everipedia (2nd nomination)	1915	14	2
Skycoin	1762	23	1
Factom	1665	10	1
PIVX	1063	11	1
ARDOR	958	17	1
Cardano (platform)	928	7	1
OmiseGO	887	9	1
DigiByte	526	8	1

Information on particular cryptocurrency can be created in each language version of Wikipedia independently and each language community can decide if the particular language version of the article can be deleted.

#### 4.1.3. Analysis Procedure

As mentioned in the previous section, a vast range of measures and indicators [33,41] can result from the processing and analysis of administrative data and Wikipedia knowledge. The careful observation of these indicators allows the formulation of judgments about certain aspects of Wikipedia performance in terms of a major knowledge provider on the Internet.

An important aspect that has not been considered in the research so far is the objectivity of the knowledge presented in Wikipedia articles. This dimension is similar to assessing the quality of article content but requires a different approach. In particular, it involves a set of actions that differs from that of the basic assessment of the quality of the articles.

It should be noted that the objectivity of the statements made in the articles is particularly important in the case of content such as the description of cryptocurrencies or independent payment systems. This is because the presentation of such content affects the opinions of potential users of such solutions. The turnover and use of cryptocurrencies are related to their relative popularity. Relative popularity should be understood as the preferences of participants of the settlement system of one platform over others. On the other hand, the popularity of a settlement platform is often correlated with the market results of the related cryptocurrency. Hence the exceptional motivation of people connected with particular platforms to promote these platforms' positive image. This is followed by the need to disseminate favorable information about a particular cryptocurrency. Such a scheme of operation is in the interest of the participants performing various roles related to the cryptocurrency circulation. These roles include, among others, creators, owners, extractors or propagators. Wikipedia is an ideal place to present knowledge about cryptocurrency systems. This is supported by its independent and impartial character. Consequently, the adversaries of a concurrent crypto coin may make attempts to spread false or discouraging knowledge to undermine the trust of the article readers. An example that is related to the cryptocurrency market is the article about a Swiss company (see [42]).

Several specific steps were executed to perform the sentiment analysis. These steps were planned and designed carefully to obtain a fully correct measurement outcome. The process was in line with approaches taken from a broad set of literature published on the sentiment analysis of microblogs [43] which is a similar case to the one examined herein.

First of all, the objects to be examined were identified. Although, as mentioned, there exist 45 articles in the English Wikipedia that aims to cover specific cryptocurrency systems in the reported sentiment analysis a limited number of 43 of them were taken into account. The limitation was the result of the initial qualification based on the quality of the articles. Hence, the final number of examined items was the effect of the filtering process.

This filtering was the second step in the research procedure. The cryptocurrency category contains a set of articles whose average quality for English edition is 44.4 (see Table 11). The quality was

assessed with the WikiRank indicator discussed in the previous section. However, one can identify articles with low quality which often have little content, or which are stubs or redirections to other entries. During the filtering stage, all the articles that were considered to be of extremely poor quality were removed from the list of the research objects. The rationale for this is that the sentiment detection algorithms as an NLP operation demand a minimum length of text sample to produce any meaningful result. The processed text sample itself has to be readable and coherent.

The third step was the text extraction from the particular Wikipedia entries that was included in the research set. In general, the extraction process was analogical to the techniques described in Section 3.2.

The extracted set of articles contents was cleaned using special text filters. It was necessary because the raw article texts contain rich additions in the form of both machine-readable (formatting and patterns) as well as human-readable (references and footnotes) passages.

After the steps that constituted the collection and cleaning of data, the actual analysis procedure could commence.

The TextBlob Python library [44] was used to process the collection of texts. The pattern-based model was chosen as it is a default algorithm of the library. This model uses the resources provided by the Pattern library [45]. It does not require any training preprocessing as the language resources are already contained within the library. It is also worth noting that while TextBlob provides only English sentiment analysis processor, the underneath Pattern library allows performing such analysis also in French and Dutch. Each article was split into a collection of separate sentences. These single sentences were the main unit of analysis. Each sentence from the article was assigned a sentiment score using the tool provided by the TextBlob library. This score indicated if the polarity of the sentence, i.e., whether it is positive, neutral or negative with values from +1 to -1, respectively. The overall result for the whole article was calculated as an average score of sentences that are the part of the article.

Table 5 contains the results of conducted sentiment analysis for the whole set of considered articles. This is a joint table where columns 3 and 4 are the continuation of columns 1 and 2. The first column presents the name of the Wikipedia entry which is usually related to the name of the relevant crypto coin. The second column shows the sentiment analysis outcome for any particular entry.

**Table 5.** Sentiment analysis results for English Wikipedia articles on cryptocurrencies.

Wikipedia Article	Sentiment	Wikipedia Article	Sentiment
Counterparty (platform)	14.50%	Libra (cryptocurrency)	4.20%
Steem	11.16%	Power Ledger	4.14%
Dogecoin	11.02%	Estcoin	4.00%
Stellar (payment network)	10.39%	Ethereum	3.15%
Royal Mint Gold	9.62%	Primecoin	3.10%
Bitnation	9.17%	Zerocoin protocol	3.07%
Bitconnect	8.70%	Auroracoin	2.83%
PotCoin	8.36%	MazaCoin	2.78%
EOS, IO	7.42%	Bitcoin Cash	2.39%
Coinye	6.62%	KodakCoin	2.22%
Litecoin	6.30%	Nxt	2.01%
Vertcoin	6.28%	Tether (cryptocurrency)	2.01%
Spintria	6.23%	NEM (cryptocurrency)	1.61%
Initiative Q	6.05%	Filecoin	1.30%
Dash (cryptocurrency)	5.79%	Monero (cryptocurrency)	0.67%
Peercoin	5.68%	Petro (cryptocurrency)	0.51%
Ethereum Classic	5.64%	Ripple (payment protocol)	0.50%
Tezos	5.60%	Bitcoin Gold	-1.70%
Gridcoin	5.00%	NEO (cryptocurrency)	-1.91%
Bitcoin	4.91%	Verge (cryptocurrency)	-3.13%
Bitcoin	4.90%	Zcash	-3.30%
Namecoin	4.76%		

As mentioned above the procedure of the analysis was the same for each entry therefore the results are fully comparable. It allows the creation of a simple ranking that provides information about the position of an article relative to other articles. The single percent value indicates the bias of the text towards a positive description of particular cryptocurrency or a negative tendency of the editors. The positive inclination is indicated by the values that are larger than 0. Consequently, the negative attitude encoded within the text yields the result below 0.

Interestingly, almost all of the articles have a rather positive tendency. The Ripple is the last entry with this kind of bias. Although, the mentioned payment protocol is the closest (0.50%) to the neutral value of 0. It means that the text that describes this crypto coin is almost well-balanced. Continuing, only 4 entries in the ranking are among articles with negative sentiment. As can be seen, the Bitcoin which can be regarded as a de-facto standard among cryptocurrencies is presented with a moderately positive attitude (4.90%). One needs to remember that the Bitcoin article is specific because of its superior quality which will be discussed in the next section. On both ends of the ranking, there are Counterparty platform (14.50%) with the most positive result at all and Zcash with the most negative result (−3.30%). What is also striking is that the average sentiment result is at the level of 5% which is clearly under the neutrality level of 0.

#### 4.1.4. Quality vs. Bias Ranking

Wikipedia from its root is a collaborative project of knowledge base creation in an encyclopedic form. It is an open and quasi-democratic system in which anyone can be a reader or creator. Nevertheless, the creators must abide by the rules. These rules are elaborated by the Wikipedia community. It means that not all content is considered to be appropriate to be presented on the portal. This is important both in terms of article subjects as well as the text within articles itself. Similarly, to other encyclopedias the topics of the entries ought to be prominent enough and unquestionable. Furthermore, to meet the editorial requirements, they should only contain provable and source-based statements. The subjects related to the crypto-economy are relatively fresh therefore they may be considered by the part of the community as insufficiently grounded. Moreover, the motives of the authors of articles and text fragments may be interpreted in diverse ways. This interpretation may include the economic motivation which was mentioned in the previous section. That is why the number of articles that belong to the crypto coin domain is constrained.

In Section 3.2 the WikiRank measure was mentioned. It is a synthetic numeric tool that permits the valuation and comparison of Wikipedia articles' quality. It can be used regardless of the Wikipedia version as it is language-agnostic. The value of the measure is normalized in the range of 0–100. It is important to understand that the WikiRank is a complex instrument. It means that its value is the resultant of many factors related to the assessed article. It is the result of the analysis of over 100 features of any Wikipedia entry. However, it is an example of a shallow analysis which means that the WikiRank does not take into account the semantics of the article. On the contrary, the presented ranking of the sentiment demonstrated in Section 4.1.3 is the case of a simple semantic analysis. It means that it focuses on certain aspects of the meaning of the phrases that make up the article. Thus, the results are not bound to syntactic or other formal characteristics.

There are three rankings presented in Table 6. The list of the cryptocurrencies descriptions in English Wikipedia is given with the corresponding values of their quality scores. The presented data are extended with the sentiment ranking data. The combination of these values is a good opportunity for making a comparison. It also makes an analytical space in which two factors are assembled. The quality score is the syntactic dimension whereas sentiment is the semantic dimension. The quality score demonstrates the formal criteria that are responsible for the excellence of the article. The semantic dimension at least partially determines the level of objectivity of the information contained in the article text. Additionally, information about the number of authors of the entries is given. This is the cumulative number of authors that contributed to the article text for the whole lifespan of its existence.



**Table 6.** Comparison of WikiRank quality score-based ranking, sentiment assessment and number of authors for selected cryptocurrency articles in English Wikipedia.

Name	Quality Score	Sentiment	Number of Authors
Bitcoin	96.33	4.90%	2 947
Ripple (payment protocol)	42.61	0.50%	430
Ethereum	78.72	3.15%	680
Dogecoin	58.56	11.02%	646
Namecoin	48.92	4.76%	190
Peercoin	32.05	5.68%	176
Petro (cryptocurrency)	67.96	0.51%	134
Litecoin	34.27	6.30%	514
Bitcoin Cash	67.77	2.39%	273
Monero (cryptocurrency)	59.73	0.67%	263
Dash (cryptocurrency)	46.67	5.79%	215
Zcash	40.96	-3.30%	126
AVERAGE	56.21	3.53%	549.5

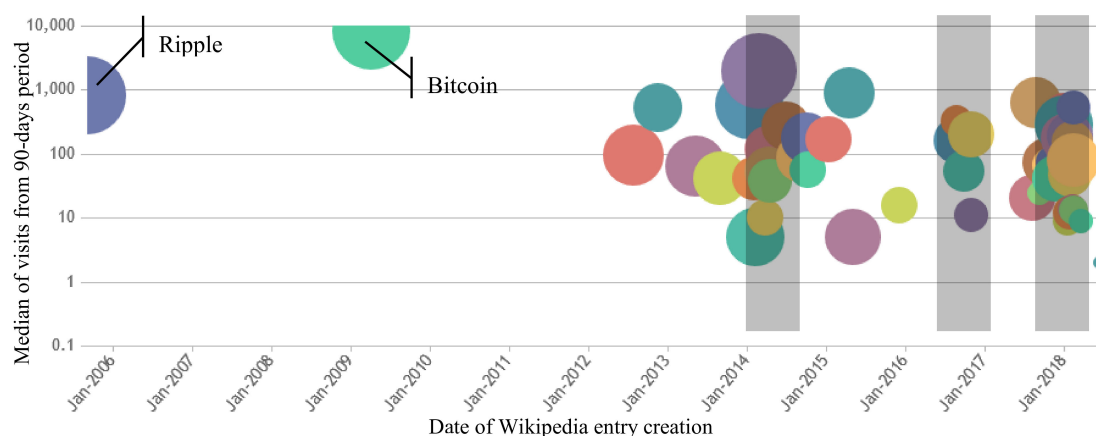
The presented data are related to the English version of Wikipedia only. It is because, in contrast to the quality measure, the sentiment measure is heavily dependent on the language of the analysis. Therefore, different instrument (language resources) is needed to perform analysis for another language.

As can be seen in the table the values of both measures are not correlated. Pearson's correlation coefficient is at the level of  $-0.061$ . It means that the editors' attitude towards particular cryptocurrencies is completely independent of the editorial level and completeness of the description. The article about Bitcoin is the unquestionable quality leader (96 points). It is the aftermath of Bitcoin's history as the first cryptocurrency as well as almost the earliest description in Wikipedia within the category. It has also a moderate level of emotional charge which is close to the average sentiment of all entries that deal with the crypto coins. Nevertheless, looking at the top of the quality ranking the sentiment level can be very diversified. The second-best article that describes Ripple has the least emotional entry. It is almost neutral in its description. The third-place—Ethereum—has results close to the Bitcoin in both dimensions. But the Dogecoin which is in the fourth place according to the quality is exceptional because its description is one of the most positive from all the categories. It is ranked third when it comes to a positive level of sentiment. The most negative article which is about Zcoin has a quality level almost perfectly equal to the average quality of all articles.

#### 4.2. Cryptocurrencies Popularity Model

This part of the article contains details about a basic model of cryptocurrencies popularity. The model permits the assessment of the demand for any Wikipedia article that covers the information on a particular crypto coin. The data for the model is taken from the Wikipedia statistics stream. As mentioned earlier, the necessary set of data are public and available by processing methods presented in Section 3.2.

In Figure 2 one can see the popularity of pages that describe various cryptocurrencies. The horizontal timeline represents the moment of Wikipedia entry creation. The points on the vertical logarithmic scale represent the median of site visits counted for the last 90 days as of the beginning of 2019. Bitcoin is an unquestionable leader whereas surprisingly it is not the earliest entry. The earliest added entry into English Wikipedia is Ripple payment protocol which was not treated as a cryptocurrency at the moment of creation. The spot sizes are correlated with a market value of cryptocurrency or token.



**Figure 2.** Popularity of cryptocurrencies on English (EN) Wikipedia and the timeline of the article creation events.

#### 4.2.1. Popularity in Google Trends

Google Trends is a tool for analyzing social interest in a given subject. This analysis has a primarily spatial and temporal dimension. The tool uses search queries data and displayed results of the largest web search engine from Google.

It is a particularly useful instrument to measure and compare the strength of interest in specific topics on the Internet. Its main drawback is that the full information on the trends is not disclosed to the user. The real figures are not presented. Instead, only a relative index is given. The values of the index are subject to change depending on the analysis context. The revealed data are enough to make some basic comparisons as well as for particular analytical tasks. On the other hand, they do not give the full picture of the use of the Internet. It also provides data with variant granularity depending on the time range. The granularity rules are described for instance in [46]. Another disadvantage, which was already mentioned is that it represents a single and specific source of trend data.

Several tests were conducted related to the context of cryptocurrencies on the relationship between Google Trends and the popularity of visited entries on Wikipedia. The results of these tests are promising but ambiguous.

The study examined the correlation between the time-series describing trends in 15 countries and the visits to Wikipedia articles in 27 language versions. In particular, some of the language versions corresponded to the countries for which the Google Trends indications were studied. The research was conducted for numerous crypto coins. But four of them, namely Bitcoin, Ethereum, Litecoin, Ripple are the key coins on the market. Related data are shown in Table 7.

In this part of the study, the courses defined in two-time resolutions were analyzed: with daily and weekly data.

As mentioned, the correlation of time-series with weekly granulation was also analyzed (see Table 8). In this group of collected data, the correlation between Google Trends and visits to Wikipedia cryptocurrency-related articles was much greater. This result is believed to be due to two reasons. First, lower time resolution makes the data smoother. Second, in this case, it was not necessary to transform the original data from Google. As a result, these time series are more reliable. The transformation is possible for example with the algorithm taken from [47].

**Table 7.** Correlations between Wikipedia all agents visits and Google Trends weekly data.

Country	Bitcoin	Ethereum	Litecoin	Ripple
World	0.908	0.816	0.950	0.957
Australia	0.924	0.796	0.947	0.939
Brazil	0.688	0.803	0.900	0.933
Belarus	0.855	0.707	0.755	0.933
China	0.774	-	0.799	0.017
Germany	0.816	0.755	0.923	0.925
Egypt	0.951	-	0.068	0.809
Spain	0.777	0.817	0.911	-
France	0.931	0.744	0.951	0.928
Great Britain	0.917	0.794	0.934	0.937
India	0.953	0.797	0.937	0.926
Iran	0.905	0.422	0.809	0.783
Italy	0.894	0.754	0.933	0.926
Japan	0.871	0.792	0.920	0.896
Poland	0.921	0.044	0.905	0.082
Portugal	0.694	0.633	0.912	0.908
Russia	0.823	0.790	0.909	0.966
United States	0.921	0.825	0.950	0.971
Venezuela	0.681	0.414	0.587	-

**Table 8.** Correlations between other user types of Wikipedia visits and Google Trends weekly data.

Country	Desktop User	Mobile App User	Mobile Web User	Desktop User	Mobile App User	Mobile Web User	Desktop User	Mobile App User	Mobile Web User
	Bitcoin			Ethereum			Litecoin		
World	0.887	0.879	0.904	0.712	0.838	0.861	0.913	0.955	0.961
Australia	0.903	0.890	0.920	0.684	0.835	0.861	0.910	0.953	0.959
Brazil	0.401	0.702	0.897	0.782	0.687	0.744	0.842	0.567	0.934
Belarus	0.823	0.854	0.865	0.695	0.727	0.718	0.737	0.694	0.766
China	0.751	0.588	0.799	-	-	-	0.790	0.689	0.790
Germany	0.771	0.813	0.835	0.730	0.700	0.784	0.915	0.922	0.925
Egypt	0.919	0.883	0.955	-	-	-	0.072	0.079	0.048
Spain	0.736	0.739	0.794	0.794	0.792	0.835	0.905	0.864	0.913
France	0.891	0.899	0.946	0.716	0.768	0.772	0.941	0.912	0.956
UK	0.895	0.891	0.914	0.684	0.824	0.856	0.896	0.937	0.945
India	0.938	0.920	0.950	0.681	0.846	0.869	0.898	0.949	0.954
Iran	0.780	0.698	0.928	0.431	0.365	0.404	0.743	0.543	0.813
Italy	0.857	0.793	0.908	0.746	0.714	0.742	0.929	0.898	0.924
Japan	0.830	0.828	0.885	0.760	0.725	0.800	0.869	0.723	0.942
Poland	0.866	0.914	0.948	0.040	0.045	0.034	0.875	0.786	0.934
Portugal	0.455	0.744	0.835	0.616	0.540	0.580	0.873	0.568	0.931
Russia	0.790	0.825	0.833	0.784	0.816	0.788	0.888	0.823	0.923
USA	0.897	0.907	0.919	0.710	0.872	0.896	0.913	0.955	0.961
Venezuela	0.647	0.616	0.694	0.400	0.322	0.417	0.571	0.440	0.591

Daily time series are much less correlated. This is because Google Trends for the analysis period longer than one month provides aggregated data only. Therefore, obtaining daily data for a sufficiently long analysis period requires the data to be collected manually for each month and then converted accordingly. This transformation consists in finding a common base and scaling the daily data from the following months. However, this procedure is burdened with a certain level of error due to incomplete information at the input of this process. This is a consequence of one of the GT disadvantages mentioned above.

#### 4.2.2. Modeling Popularity

The model is based on a number of assumptions that concern the relation between the default schemes of Wikipedia and cryptocurrency usage. All the assumptions must be met so that the model gives appropriate outcomes. These assumptions include:

- proportionality of the number of visits and popularity in the tested period,
- proportionality of division of visits between users of different nationalities that use the same language,
- the population of a given nationality use, in general, their national language (the rest that so not obey to this rule are negligible),
- the ratio of visits of users from given territory to all visits of the certain language version is constant in time.

Following equation formulates the model:

$$P(c)_t = \sum_{lang=1}^N \alpha(t)_{lang} * V(c)_{lang} \tag{1}$$

The symbols in the model represent:

- $P$  is the popularity estimate of certain cryptocurrency  $c$  in a given territory (country)  $t$ ,
- $\alpha$  represents the coefficient indicating the ratio of visits of users from country  $t$  to all visits of the certain Wikipedia language version,
- $V$  is the overall number of monthly views of selected cryptocurrency article in a given Wikipedia language version,
- $N$  is the total number of analyzed Wikipedia internationalized editions.

The data from all Wikipedia language versions were processed with the model. Likewise, all the cryptocurrency-related articles were the subject of the analysis. Table 9 shows selected values of  $\alpha$  coefficient for particular combinations of states and Wikipedia national versions. For example, the fifth row (DE) and second column (de) shows that 79% of users form Germany reads articles in German Wikipedia. The second most frequently langue used by the same population is Persian. However, articles in this language are only read by 4% of users living in Germany.

**Table 9.** Ratio of visits of users from selected countries to all visits of the certain Wikipedia language version.

Country	ar	de	en	es	fa	fr	it	ja	pl	pt	ru	zh
AU	0.0012	0.0006	0.0272	0.0005	0.0024	0.0007	0.0008	0.0008	0.0005	0.0006	0.0007	0.0111
BR	0.0014	0.0006	0.0065	0.0032	0.0003	0.0018	0.0014	0.0003	0.0003	0.7294	0.0011	0.0005
BY	0.0001	0.0001	0.0004	0.0	0.0001	0.0001	0.0001	0.0	0.0005	0.0	0.0395	0.0001
CN	0.0	0.0	0.0001	0.0	0.0	0.0	0.0	0.0001	0.0	0.0	0.0	0.0078
DE	0.0153	0.7896	0.0251	0.0059	0.0399	0.0103	0.0115	0.0026	0.0192	0.0049	0.0171	0.0101
EG	0.1282	0.0003	0.0016	0.0001	0.0001	0.0004	0.0002	0.0001	0.0003	0.0001	0.0005	0.0002
ES	0.0017	0.0027	0.0057	0.2231	0.0019	0.0046	0.0042	0.0003	0.0015	0.0024	0.002	0.0021
FR	0.0077	0.0054	0.0106	0.0048	0.0127	0.6652	0.0073	0.0013	0.0045	0.0036	0.0067	0.0066
GB	0.0081	0.0038	0.0989	0.0025	0.0215	0.0056	0.0067	0.0014	0.0177	0.0034	0.0044	0.0154
IN	0.0024	0.0006	0.0758	0.0019	0.001	0.0017	0.0005	0.0004	0.0006	0.0008	0.002	0.0018
IR	0.0021	0.0001	0.0153	0.0001	0.7557	0.0002	0.0001	0.0	0.0	0.0	0.0001	0.0001
IT	0.0018	0.0041	0.0077	0.0026	0.0027	0.0039	0.9175	0.0004	0.0018	0.0015	0.0026	0.0025
JP	0.0005	0.0009	0.005	0.0005	0.0023	0.0014	0.0007	0.9631	0.0004	0.0015	0.0011	0.0304
PL	0.0007	0.0024	0.0059	0.0007	0.0009	0.0011	0.0014	0.0002	0.9055	0.0006	0.0054	0.0006
PT	0.0001	0.0005	0.0022	0.0007	0.0001	0.0011	0.0008	0.0	0.0003	0.0961	0.0005	0.0003
RU	0.0022	0.0021	0.0069	0.0016	0.0023	0.002	0.0012	0.0007	0.0016	0.0015	0.6102	0.0023
US	0.1383	0.0116	0.4391	0.0829	0.094	0.0565	0.012	0.0112	0.0119	0.1174	0.0208	0.1046
VE	0.0001	0.0	0.0004	0.0343	0.0	0.0001	0.0001	0.0	0.0	0.0001	0.0001	0.0

One can calculate the assessment of cryptocurrency popularity  $P(c)_t$  based on the Wikipedia articles' views. This popularity may be presented as a nominal value of the total visits within a given period. Alternatively, it may take the form of the fraction of total visits divided by the size of the country's population that has internet access. The exemplary fractional results are given in Table 10.

**Table 10.** Pageviews of four most important cryptocurrencies for selected countries over the world.

	Bitcoin	Ethereum	Litecoin	Ripple
AU	0.002676%	0.000706%	0.000157%	0.000284%
BR	0.004212%	0.000656%	0.000252%	0.000317%
BY	0.004500%	0.001808%	0.000307%	0.000472%
CN	0.005048%	0.001432%	0.000293%	0.000248%
DE	0.004259%	0.000839%	0.000205%	0.000412%
EG	0.006144%	0.000464%	0.000307%	0.000251%
ES	0.003673%	0.000823%	0.000250%	0.000276%
FR	0.004505%	0.000873%	0.000195%	0.000388%
GB	0.002693%	0.000696%	0.000157%	0.000285%
IN	0.002584%	0.000684%	0.000153%	0.000284%
IR	0.008082%	0.000724%	0.000234%	0.000441%
IT	0.003446%	0.000656%	0.000158%	0.000265%
JP	0.001288%	0.000379%	0.000055%	0.000218%
PL	0.004669%	0.000833%	0.000257%	0.000196%
PT	0.003871%	0.000675%	0.000230%	0.000312%
RU	0.004454%	0.001783%	0.000304%	0.000467%
US	0.002772%	0.000701%	0.000162%	0.000286%
VE	0.003736%	0.000832%	0.000261%	0.000272%

As can be seen, from the point of view of the share of internet users Bitcoin is extremely popular in Iran (0.008082%). Whereas, Ethereum is the most popular in Belarus. Such relative popularity of other crypto coins in Belarus is also true for Litecoin (first place together with Egypt) and Ripple (the result that is comparably high with Russia).

#### 4.2.3. Model Evaluation

The model was estimated by comparison of its output with data acquired from Search Engine Optimization (SEO) tools. The popularity indicators of particular cryptocurrency-related websites from such sources as HypeStat [48], Alexa [49] and others. Based on the obtained data three rankings were created. The ranking determines the relative position of different countries in terms of the information demand for a particular cryptocurrency. Similar rankings were used earlier by researchers [50]. The comparison of all the sequences can be made with the calculation of Spearman's rank coefficient. The coefficient indicates the extent of co-occurrence between the two rankings. The sequences for the Bitcoin article, popularity data of bitcoin.org website based on SEO tools and Google Trends data are depicted in Figure 3.

In the picture, one can see three sequences of country codes. The numbers represent their position within the sequences. For example, according to Google Trends Ireland (IR) is on 3rd place. At the same time, this country is situated on the 2nd place according to Wikipedia. Another example—Russia (RU) is on 10th place in both rankings—based on Google Trends and Wikipedia. We can also observe some similarities between rankings based on Wikipedia and Internet statistics: Turkey (TR) takes 3rd place and Thailand (TH) is on 14th place in both of the rankings. Egypt (EG) took 5th and 7th place in Wikipedia and Internet ranking, respectively.

Spearman's rank coefficient was used to estimate the relevance of the connection of Wikipedia and Internet ranking orderings. The value of the coefficient  $r_s$  was at the level of 0.56235 whereas  $p$  (2-tailed) = 0.00029. Such results indicate that by usual standards the association between the values of analyzed two datasets is statistically significant.



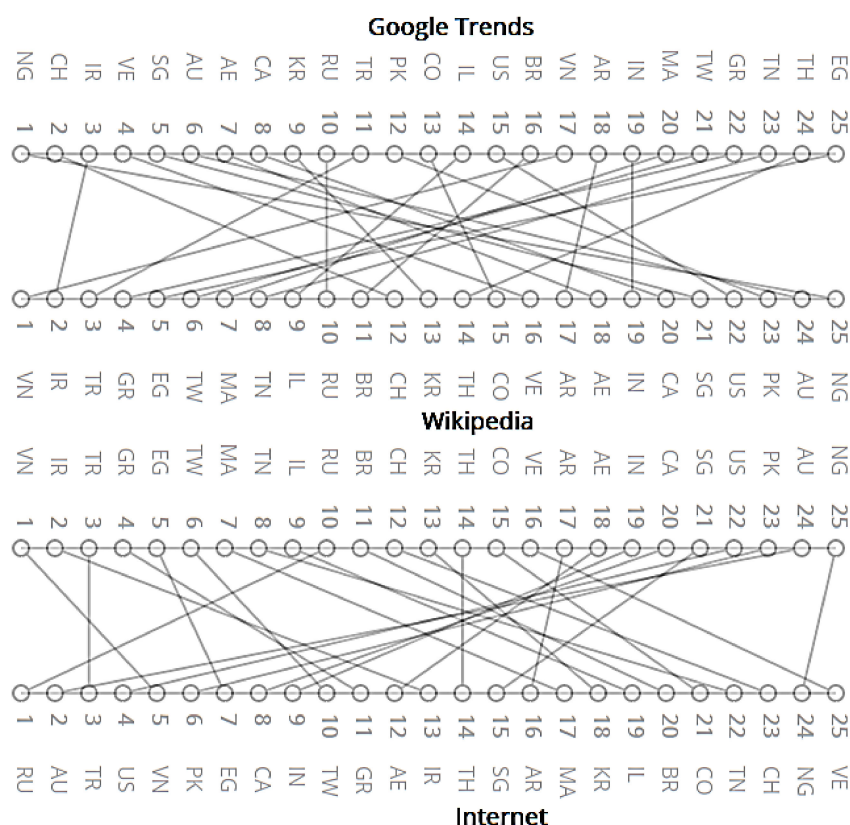


Figure 3. Comparison of Wikipedia, Google Trends and Internet rankings of Bitcoin visitors.

### 4.3. Blockchain Regulations

The idea of electronic cash existed before the advent of Bitcoin. But it was the first cryptocurrency that allowed peer-to-peer exchange and so was fully third-party independent. The lack of a body of authority and control raises questions about the trustworthiness and reliability of the technology. Consequently, many states and public institutions in numerous countries expressed their doubts [51] or even warned the potential users of the crypto coins. National regulations and attitudes towards cryptocurrencies are varied. This issue is raised in the literature [52,53]. There exist also an article in Wikipedia that tackles the legal status of crypto coins from the point of view of individual legislatures.

The regulation of technological advancement is a sophisticated issue [54]. It is a problem of properly shaped and balanced policy-making strategy. It also needs legislative bodies that are objective in judgments as well as well-acquainted with details of how the new solution works. The task becomes even more complicated when it comes to the technology that is by its nature hard to control. It is because this kind of setting cannot be supervised or steered easily.

Blockchain in general and crypto coins as its main application is such kind of setting. As most blockchain architectures are largely decentralized. It means that they are highly resilient and tough to tackle with the hitherto means. On the other hand, the vision spread by some individuals of disruption revolutionary potential and other fears of the technology fuels the anticipations of certain stakeholder groups to curb the unlimited introduction of the blockchain into many spheres of economic life. It is also worth mentioning that even the blockchain industry itself expects to certain degree introduction of some limitations, especially for adhering to the actual state of affairs to a legal framework.

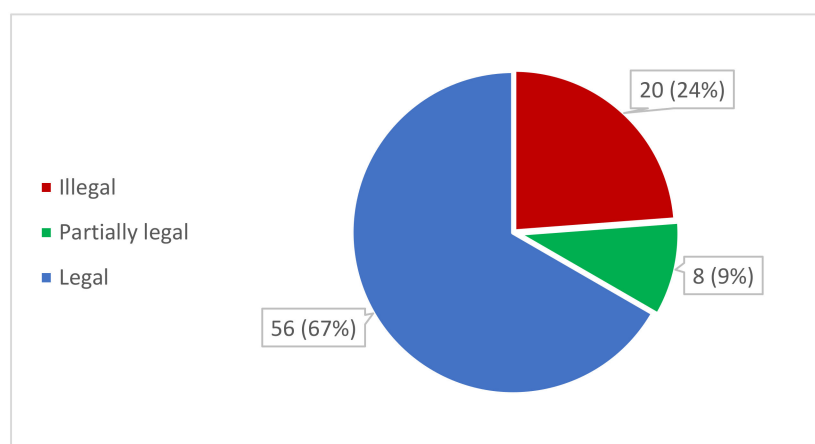
In this complicated landscape, the states all over the world reside in a rather chaotic condition. It is both too early and too late to force the technology off. In addition, as the Internet is transnational it is almost impossible to ban cryptocurrencies at all.

Different countries take different steps. Initially, the technology was often prohibited. Another approach was to criticize and defame it. Some more open communities decided to wait in order to observe how the technology would evolve by itself. Although even in this case some institutions made discouraging messages to the public.

After a decade has passed since the introduction of Bitcoin the status of cryptocurrencies is very diversified across the globe [53,55]. In a very broad way, one may divide the states on those that ban the usage of the technology and those that allow for it. Certainly, when analyzing more scrupulously the situation is more sophisticated. Some states permit the use by private citizens but disallow making transactions or other deeds by institutions (subject limitation). Some countries limit specific types of activity concerning cryptocurrencies with exchanges being the most prominent example (action limitation). Another way of influencing the crypto coins industry is by shaping the specific rules and methods of taxation [56]. The construction of the tax system is oftentimes an indicator of the general policy of society towards the selected economic phenomenon. It can be tailored to incentivize certain activities or adversely, it can be thought in a way to limit particular behaviors.

The classification of the status of the states' governments' attitude toward cryptocurrency technologies is not a straightforward task. There are several reasons why this is the case. First of all, the comparison of the legislative framework is not always possible. Second, some sources bring contradictory information. Third, the state of the regulations change in time. Finally, there are often many institutions and governing bodies that have fragmented or overlapping prerogatives.

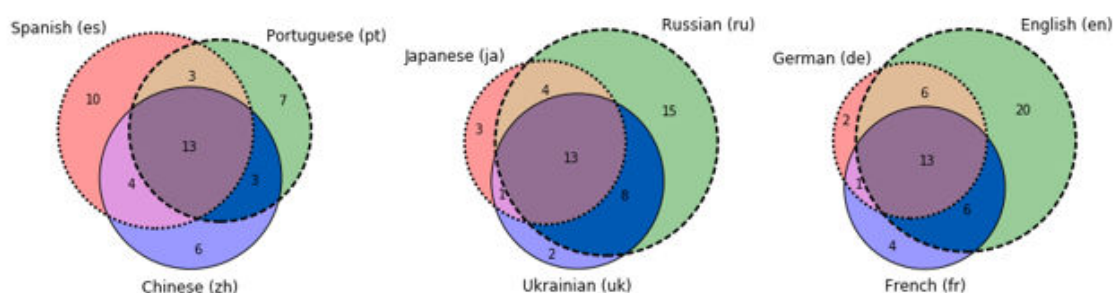
For the purpose of this study, numerous sources were examined [52,57]. Wikipedia article about the legality of Bitcoin [51] was chosen as a basis for preparing dataset on the subject. In the end, a simple classification of countries was prepared. This three-step grouping recognizes the following statuses: illegal, partially legal and legal. The two extreme classes are self-explanatory. The middle class contains states with mixed approaches ranging from these which changed their status from illegal to countries which disallow some form of activities at least for certain groups of subjects. The division of classes is shown in Figure 4.



**Figure 4.** Ratio of jurisdictions in which Bitcoin (and other crypto coins) are permitted or forbidden to use.

#### 4.3.1. Confronting Popularity vs. Legality

Some of the cryptocurrencies are presented in Wikipedia in a limited number of languages. Figure 5 shows the coverages between Wikipedia languages in articles about cryptocurrencies.



**Figure 5.** Venn diagram of sets of cryptocurrency-related articles in selected editions of Wikipedia. See <http://data.lewoniewski.info/crypto/> for interactive version.

Having information about the language versions of a particular cryptocurrency, one can calculate the general popularity of all described cryptocurrencies in each language. Table 11 shows language versions of Wikipedia with the highest number of articles about cryptocurrencies with the aggregated number of page views as well as their medians.

**Table 11.** Basic facts about cryptocurrencies description in Wikipedia language editions.

Language	Num. of Articles 2020 (2018)	Average Quality	Total Visits	Visits Median
en	45 (39)	44.4	239,259	8580
ru	40 (38)	35.0	48,925	1655
zh	26 (27)	39.3	23,918	847
pt	26 (25)	36.1	8102	290
es	30 (24)	36.4	22,756	796
uk	24 (22)	37.9	1878	62
fr	24 (21)	35.2	17,422	608
it	19 (20)	29.0	9891	347
ja	21 (19)	32.0	13,291	478
de	22 (18)	43.3	37,803	1395

In this section, the output data from the model introduced above are applied to the dataset describing the attitude towards the cryptocurrency technology of particular countries. The process aims to check whether there is any relationship between cryptocurrencies’ popularity in given states and the governmental approach that allows or disallows of the technology being used.

In order to identify the relation, a predictive model based on data mining techniques was decided to be built. The production of the model which will be able to produce reasonable results would be identical with the existence of a cause-and-effect link between the independent and dependent variables of the model.

Before starting this study two possible hypotheses were plausible. H0 stating that “the popularity of the cryptocurrency technology simply reflects the openness of the governing bodies and the general economic freedom”. The alternative hypothesis H1 is that “contrary to the official regulations users in countries with less economic freedom are more interested in the use of alternative payment methods than the official money because of their limited trust to their countries’ economies (and overregulation)”.

The Random Forest algorithm was used to build the classification model using Weka software [58]. The model had two independent variables: the estimated Wikipedia popularity in both integral and percentile forms. The dependent variable was binary—the status could be either legal or illegal. The model was generated using 100 trees and 10-fold cross-validation on the training set. The number of observation instances (countries) totaled 76.

In the end, the received model can be characterized by the following quality characteristics: Precision of 66%, Recall at the level of 67%. The F-Measure equaled 67%.

The obtained results indicate a moderate link between the allowance of the use of cryptocurrencies and their popularity in particular countries. This confirms the principal hypothesis H0 and abolishes its alternative H1. The model was fed with almost half of the world's countries data. Especially all the larger nations were considered. It can be treated as an indication for the governing bodies when they consider the strategies and their outcomes of regulating cryptocurrencies. The results have also an important impact on analyses related to the cryptocurrency markets. It indicates that the level of regulations and especially their changes will be reflected in the volume and transactions performed in a particular country.

## 5. Implications

The presented results are especially vital for the theory-oriented audience as they indicate the socio-economic relations related to the information demand in the crypto-economy domain. Moreover, the paper reveals insight for some non-trivial issues and answers theory-grounded questions. They are also important for practical use. They should be of interest both for the Wikipedia editors and content creators as well as for the marketers and developers of the cryptocurrencies. For the first group of recipients, the given answers should indicate the situation and challenges related to the edition of particular areas of the open encyclopedia. The latter group can get aware of the global and local trends and competitive advantages of particular digital payment systems.

In the Introduction section a number of research questions (RQs) were posted. These were the questions that drove the authors to conduct the presented studies and their results. Within the Implications section these questions will be synthetically answered below.

In general, the emotional balance of Wikipedia articles that describe cryptocurrencies is on reasonable level with the tendency toward neutrality (RQ1). However, it does not mean a perfect balance. The positively marked articles prevails with Bitcoin being not only the standard for the crypto coins themselves but also a standard for the encyclopedia entry. Making the Bitcoin article almost in the center of the sentiment ranking. The possible explanation is that authors of other articles modeled their texts on the one preliminary and most fundamental to the domain.

There is no simple association between sentiment indication and quality of an article (RQ2). The possible interpretation of the combination of the two measures is that the quality can be treated as a confidence level of the measured sentiment. One of the factors that influence WikiRank is the length of the article text. In addition, the longer is the assessed text by the sentiment analysis algorithm the more accurate its indication should be. Likewise, there has not been detected a relation between the number of authors and the sentiment level.

There exist a strong similarity of patterns between Google Trends users' interests—which reflects a major popular search engine statistic—and the number of visits in Wikipedia (RQ3). The levels of correlation coefficients are usually high or very high across different cryptocurrencies and languages with only individual cases of significant lack of compliance.

The model proposed in Section 4.2. Its evaluation presented in Section 4.2.3 shows that it is possible to a certain degree to track the cryptocurrencies popularity on a geographical basis (RQ4). The demand for the Wikipedia information allow modeling the popularity of crypto assets described in articles and particular language versions. The particular way of preparing the model is given in the mentioned sections.

Similarly, the model presented in Section 4.3.1 demonstrate how one can bind the national attitudes towards the crypto economy with the popularity of cryptocurrencies in particular countries (RQ5). The evaluation of the model tells that, although the model results are not fully satisfactory, the binding is possible. The obtained results support the hypothesis that the popularity of the cryptocurrency technology in countries generally depends on the openness of the governing bodies and the economic freedom. It means that the less restrictions are imposed the more popular is the usage of the technology.

Finally, there are several factors that influence the number of cryptocurrencies described in Wikipedia (RQ6). The important factor is the popularity of the crypto asset to be described. Nevertheless,

a recurring situation is that added entries are deleted. The two most important reasons for the removal of cryptocurrency articles are suspicion of unambiguous advertising or promotion and no indication of importance. In total, over 300 articles about cryptocurrencies were deleted from the English version of Wikipedia with more than half of them using speedy deletion procedure. This is a large amount, given that there are only 45 Wikipedia articles describing individual cryptocurrencies in this language at present.

## 6. Conclusions

This study aimed to present the results of analyses and models built to investigate the perception of cryptocurrencies. The research is concentrated on the use of Wikipedia as an important knowledge base on the Internet. It is also a rich source of open data and metadata that can be processed. Wikipedia is a reliable source in comparison to other social or collaborative content creation platforms. Another source used is Google Trends.

We also demonstrated the results of sentiment analysis of the information presented on the topic of particular crypto coins in English Wikipedia. These results can be useful to create a ranking based on possible Wikipedia community bias in cryptocurrency descriptions.

It was also showed that Google Trends data are consistent with the information demand data from Wikipedia. The provided popularity model allows discovering the relative number of potential users in particular countries that show interest in given cryptocurrency. The obtained results from this model were compared with other sources.

Another model presented in our work uses the popularity data and confronts them with information about the level of permission or oppression of making transactions with crypto coins by citizens in particular legislatures. The approach of governments towards the new technology is a distinct and new dimension of economic freedom. The factor which is also important is the limitation of cryptocurrencies versus other forms of information flow censorship. This issue is also discussed in the paper. In future work we plan to extend number of the considered measures in order to improve obtained model.

Finally, an analysis of the dynamics of cryptocurrency-related content on Wikipedia was presented. It shows that there is a constant struggle between the reliability and good practices and subjective preferences and specific interests of Wikipedia contributors.

Besides development of new quality and popularity measures for the proposed models we also plan to expand the number of sources (such as Facebook, Twitter, Reddit) which can give additional possibilities for more complex analysis on cryptocurrencies perception in various countries.

**Author Contributions:** The contribution of the authors is as follow: conceptualization, P.S., W.A. and W.L.; methodology, W.A. and P.S.; software, W.L.; validation, P.S. and W.L.; formal analysis, W.L. and P.S.; investigation, W.L.; resources, W.L. and P.S.; data curation, W.L.; writing—original draft preparation, P.S.; writing—review and editing, P.S. and W.L.; visualization, W.L. and P.S.; supervision, W.A.; project administration, P.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript or in the decision to publish the results.

## References

1. Cuomo, M.T.; Tortora, D.; Giordano, A.; Festa, G.; Metallo, G.; Martinelli, E. User-generated content in the era of digital well-being: A netnographic analysis in a healthcare marketing context. *Psychol. Mark.* **2020**. [[CrossRef](#)]
2. Narangajavana Kaosiri, Y.; Callarisa Fiol, L.J.; Moliner Tena, M.A.; Rodriguez Artola, R.M.; Sanchez Garcia, J. User-generated content sources in social media: A new approach to explore tourist satisfaction. *J. Travel Res.* **2019**, *58*, 253–265. [[CrossRef](#)]



3. Timoshenko, A.; Hauser, J.R. Identifying customer needs from user-generated content. *Mark. Sci.* **2019**, *38*, 1–20. [[CrossRef](#)]
4. Blazquez, D.; Domenech, J. Big Data sources and methods for social and economic analyses. *Technol. Forecast. Soc. Chang.* **2018**, *130*, 99–113. [[CrossRef](#)]
5. Mestyán, M.; Yasseri, T.; Kertész, J. Early prediction of movie box office success based on Wikipedia activity big data. *PLoS ONE* **2013**, *8*, e71226. [[CrossRef](#)] [[PubMed](#)]
6. Moat, H.S.; Curme, C.; Avakian, A.; Kenett, D.Y.; Stanley, H.E.; Preis, T. Quantifying Wikipedia usage patterns before stock market moves. *Sci. Rep.* **2013**, *3*, 1801. [[CrossRef](#)]
7. Google Trends. Available online: <https://trends.google.com/> (accessed on 22 April 2019).
8. Stolarski, P.; Lewoniewski, W. *Wikipedia as an Information Source on Cryptocurrency Technology*. *International Conference on Business Information Systems*; Springer: Berlin, Germany, 2019; pp. 299–308.
9. Kristoufek, L. BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Sci. Rep.* **2013**, *3*, 3415. [[CrossRef](#)]
10. Dickerson, A. *Algorithmic Trading of Bitcoin Using Wikipedia and Google Search Volume*; Warwick Business School: Coventry, UK, 2018. [[CrossRef](#)]
11. ElBahrawy, A.; Alessandretti, L.; Baronchelli, A. Wikipedia and Cryptocurrencies: Interplay Between Collective Attention and Market Performance. *Front. Blockchain* **2019**, *2*. [[CrossRef](#)]
12. Rowley, J.; Johnson, F. Understanding trust formation in digital information sources: The case of Wikipedia. *J. Inf. Sci.* **2013**, *39*, 494–508. [[CrossRef](#)]
13. Menchen-Trevino, E.; Hargittai, E. Young adults' credibility assessment of Wikipedia. *Inf. Commun. Soc.* **2011**, *14*, 24–51. [[CrossRef](#)]
14. Lim, S. How and why do college students use Wikipedia? *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 2189–2202. [[CrossRef](#)]
15. Lim, S.; Simon, C. Credibility judgment and verification behavior of college students concerning Wikipedia. *First Monday* **2011**, *16*. [[CrossRef](#)]
16. Khadivi, P.; Ramakrishnan, N. Wikipedia in the tourism industry: Forecasting demand and modeling usage behavior. In Proceedings of the Twenty-Eighth IAAI Conference, Phoenix, Arizona, 12–17 February 2016.
17. Carrière-Swallow, Y.; Labbé, F. Nowcasting with Google Trends in an emerging market. *J. Forecast.* **2013**, *32*, 289–298. [[CrossRef](#)]
18. Bughin, J. Google searches and twitter mood: Nowcasting telecom sales performance. *Netnomics Econ. Res. Electron. Netw.* **2015**, *16*, 87–105. [[CrossRef](#)]
19. See-To, E.W.; Ngai, E.W. Customer reviews for demand distribution and sales nowcasting: A big data approach. *Ann. Oper. Res.* **2018**, *270*, 415–431. [[CrossRef](#)]
20. Georgoula, I.; Pourmarakis, D.; Bilanakos, C.; Sotiropoulos, D.; Giaglis, G.M. *Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices*; Athens University of Economics and Business: Athens, Greece, 2015.
21. Ciaian, P.; Rajcaniova, M.; Kancs, D. The economics of BitCoin price formation. *Appl. Econ.* **2016**, *48*, 1799–1815. [[CrossRef](#)]
22. Kristoufek, L. What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PLoS ONE* **2015**, *10*, e0123923. [[CrossRef](#)]
23. Garcia, D.; Schweitzer, F. Social signals and algorithmic trading of Bitcoin. *R. Soc. Open Sci.* **2015**, *2*, 150288. [[CrossRef](#)]
24. Steinert, L.; Herff, C. Predicting altcoin returns using social media. *PLoS ONE* **2018**, *13*, e0208119. [[CrossRef](#)]
25. Colianni, S.; Rosales, S.; Signorotti, M. Algorithmic trading of cryptocurrency based on Twitter sentiment analysis. *Cs229 Proj.* **2015**, 1–5.
26. Dulău, T.M.; Dulău, M. Cryptocurrency–Sentiment Analysis in Social Media. *Acta Marisiensis. Ser. Technol.* **2019**, *16*, 1–6. [[CrossRef](#)]
27. Karalevicius, V.; Degrande, N.; De Weerd, J. Using sentiment analysis to predict interday Bitcoin price movements. *J. Risk Financ.* **2018**, *19*, 56–75. [[CrossRef](#)]
28. Hevner, A.R. A three cycle view of design science research. *Scand. J. Inf. Syst.* **2007**, *19*, 4.
29. Prat, N.; Comyn-Wattiau, I.; Akoka, J. Artifact Evaluation in Information Systems Design-Science Research-a Holistic View. In *PACIS*; Association for Information Systems: Atlanta, GA, USA, 2014; p. 23.

30. Arnott, D.; Pervan, G. Design science in decision support systems research: An assessment using the Hevner, March, Park, and Ram Guidelines. *J. Assoc. Inf. Syst.* **2012**, *13*, 1. [CrossRef]
31. Horn, C.; Manduca, C.; Kauchak, D. Learning a lexical simplifier using wikipedia. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Baltimore, MD, USA, 22–27 June 2014; pp. 458–463.
32. Zesch, T.; Müller, C.; Gurevych, I. Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary. In Proceedings of the LREC, Miyazaki, Japan, 7–12 May 2018; Volume 8, pp. 1646–1652.
33. Lewoniewski, W. Measures for Quality Assessment of Articles and Infoboxes in Multilingual Wikipedia. In *International Conference on Business Information Systems*; Abramowicz, W., Paschke, A., Eds.; Springer International Publishing: Basel, Switzerland, 2019; pp. 619–633. [CrossRef]
34. Lewoniewski, W.; Wecel, K.; Abramowicz, W. Multilingual Ranking of Wikipedia Articles with Quality and Popularity Assessment in Different Topics. *Computers* **2019**, *8*, 60. [CrossRef]
35. Wikimedia. Available online: <https://dumps.wikimedia.org/> (accessed on 22 April 2019).
36. Erxleben, F.; Günther, M.; Krötzsch, M.; Mendez, J.; Vrandečić, D. Introducing Wikidata to the linked data web. In *International Semantic Web Conference*; Springer: Berlin, Germany, 2014; pp. 50–65.
37. Lehmann, J.; Isele, R.; Jakob, M.; Jentzsch, A.; Kontokostas, D.; Mendes, P.N.; Hellmann, S.; Morsey, M.; Van Kleef, P.; Auer, S.; et al. DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia. *Semant. Web* **2015**, *6*, 167–195. [CrossRef]
38. WikiRank. Quality and popularity assessment of Wikipedia articles. Available online: <https://wikirank.net> (accessed on 22 April 2019).
39. Wikipedia. Criteria for speedy deletion. Available online: [https://en.wikipedia.org/wiki/Wikipedia:Criteria\\_for\\_speedy\\_deletion](https://en.wikipedia.org/wiki/Wikipedia:Criteria_for_speedy_deletion) (accessed on 22 April 2019).
40. Wikipedia. Deletion policy. Available online: [https://en.wikipedia.org/wiki/Wikipedia:Deletion\\_policy](https://en.wikipedia.org/wiki/Wikipedia:Deletion_policy) (accessed on 22 April 2019).
41. Warncke-wang, M.; Cosley, D.; Riedl, J. Tell Me More: An Actionable Quality Model for Wikipedia. *WikiSym* **2013**, *2013*, 1–10. [CrossRef]
42. Wikipedia. Bitcoin Suisse. Available online: [https://en.wikipedia.org/wiki/Bitcoin\\_Suisse](https://en.wikipedia.org/wiki/Bitcoin_Suisse) (accessed on 22 April 2019).
43. Stenqvist, E.; Lönnö, J. Predicting Bitcoin price fluctuation with Twitter sentiment analysis. 2017. Available online: <http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1110776&dswid=mainwindow> (accessed on 20 April 2020).
44. Loria, S. TextBlob Python library. Available online: <https://textblob.readthedocs.io/en/dev/> (accessed on 22 April 2019).
45. De Smedt, T.; Daelemans, W. Pattern for python. *J. Mach. Learn. Res.* **2012**, *13*, 2063–2067.
46. Phillips, R.C.; Gorse, D. Cryptocurrency price drivers: Wavelet coherence analysis revisited. *PLoS ONE* **2018**, *13*, e0195200. [CrossRef]
47. Garcia, D.; Tessone, C.J.; Mavrodiev, P.; Perony, N. The digital traces of bubbles: Feedback cycles between socio-economic signals in the Bitcoin economy. *J. R. Soc. Interface* **2014**, *11*, 20140623. [CrossRef]
48. HypeStat. HypeStat Main Page. Available online: <https://hypestat.com/> (accessed on 22 April 2019).
49. Alexa. The top 500 sites on the web. Available online: <https://www.alexa.com/topsites> (accessed on 22 April 2019).
50. Najafi, I.; Kamyar, M.; Kamyar, A.; Tahmassebpour, M. Investigation of the correlation between trust and reputation in B2C e-commerce using Alexa ranking. *IEEE Access* **2017**, *5*, 12286–12292. [CrossRef]
51. Wikipedia. Legality of bitcoin by country or territory. Available online: [https://en.wikipedia.org/wiki/Legality\\_of\\_bitcoin\\_by\\_country\\_or\\_territory](https://en.wikipedia.org/wiki/Legality_of_bitcoin_by_country_or_territory) (accessed on 22 April 2019).
52. Caytas, J. Regulation of Cryptocurrencies and Initial Coin Offerings in Switzerland: Declared Vision of a Crypto Nation. *Incl. Chapter News* **2018**, *31*, 53.
53. Marian, O. A conceptual framework for the regulation of cryptocurrencies. *U. Chi. L. Rev. Dialogue* **2015**, *82*, 53.
54. Kethineni, S.; Cao, Y. The Rise in Popularity of Cryptocurrency and Associated Criminal Activity. *Int. Crim. Justice Rev.* **2019**. [CrossRef]

55. Fitzgerald, O.; Anissimov, S. Regulating Blockchain and Distributed Ledger Technologies: Challenges and Opportunities for Canadian Innovation. 2016. Available online: [https://www.cigionline.org/sites/default/files/documents/cigi\\_round\\_table\\_web.pdf](https://www.cigionline.org/sites/default/files/documents/cigi_round_table_web.pdf) (accessed on 22 April 2019).
56. Hughes, S.; Middlebrook, S. Advancing a framework for regulating cryptocurrency payments intermediaries. *Yale J. Reg* **2015**, *32*, 495.
57. Tsukerman, M. The block is hot: A survey of the state of Bitcoin regulation and suggestions for the future. *Berkeley Technol. Law J.* **2015**, *30*, 1127–1170.
58. Weka 3. Data Mining with Open Source Machine Learning Software in Java. Available online: <https://www.cs.waikato.ac.nz/ml/weka/> (accessed on 22 April 2019).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).