information

Article

Modeling Popularity and Reliability of Sources in Multilingual Wikipedia

Włodzimierz Lewoniewski *, Krzysztof Węcel and Witold Abramowicz

Department of Information Systems, Poznań University of Economics and Business, 61-875 Poznań, Poland; krzysztof.wecel@ue.poznan.pl (K.W.); witold.abramowicz@ue.poznan.pl (W.A.)

* Correspondence: wlodzimierz.lewoniewski@ue.poznan.pl

Received: 31 March 2020; Accepted: 7 May 2020; Published: 13 May 2020

Abstract: One of the most important factors impacting quality of content in Wikipedia is presence of reliable sources. By following references, readers can verify facts or find more details about described topic. A Wikipedia article can be edited independently in any of over 300 languages, even by anonymous users, therefore information about the same topic may be inconsistent. This also applies to use of references in different language versions of a particular article, so the same statement can have different sources. In this paper we analyzed over 40 million articles from the 55 most developed language versions of Wikipedia to extract information about over 200 million references and find the most popular and reliable sources. We presented 10 models for the assessment of the popularity and reliability of the sources based on analysis of meta information about the references in Wikipedia articles, page views and authors of the articles. Using DBpedia and Wikidata we automatically identified the alignment of the sources to a specific domain. Additionally, we analyzed the changes of popularity and reliability in time and identified growth leaders in each of the considered months. The results can be used for quality improvements of the content in different languages versions of Wikipedia.

Keywords: Wikipedia; reference; source; reliability; popularity; Wikidata; DBpedia

1. Introduction

Collaborative wiki services are becoming an increasingly popular source of knowledge in different countries. One of the most prominent examples of such free knowledge bases is Wikipedia. Nowadays this encyclopedia contains over 52 million articles in over 300 languages versions [1]. Articles in each language version can be created and edited even by anonymous (not registered) users. Moreover, due to the relative independence of contributors in each language, we can often encounter differences between articles about the same topic in various language versions of Wikipedia.

One of the most important elements that significantly affect the quality of information in Wikipedia is availability of a sufficient number of references to the sources. Those references can confirm facts provided in the articles. Therefore, community of the Wikipedians (editors who write and edit articles) attaches great importance to reliability of the sources. However, each language version can provide its own rules and criteria of reliability, as well as its own list of perennial sources whose use on Wikipedia are frequently discussed [2]. Moreover, this reliability criteria and list of reliable sources can change over time.

According to English Wikipedia content guidelines, information in the encyclopedia articles should be based on reliable, published sources. The word "source" in this case can have three interpretations [2]: the piece of work (e.g., a book, article, research), the creator of the work (e.g., a scientist, writer, journalist), the publisher of the work (e.g., MDPI or Springer). The term "published" is often associated with text materials in printed format or online. Information in other

format (e.g., audio, video) also can be considered as a reliable source if it was recorded or distributed by a reputable party.

The reliability of a source in Wikipedia articles depends on context. Academic and peer-reviewed publications as well as textbooks are usually the most reliable sources in Wikipedia. At the same time not all scholarly materials can meet reliability criteria: some works may be outdated or be in competition with other research in the field, or even controversial within other theories. Another popular source of Wikipedia information are well-established press agencies. News reporting from such sources is generally considered to be reliable for statements of fact [2]. However, we need to take precautions when reporting breaking-news as they can contain serious inaccuracies.

Despite the fact that Wikipedia articles must present a neutral point of view, referenced sources are not required to be neutral, unbiased, or objective. However, websites whose content is largely user-generated is generally unacceptable. Such sites may include: personal or group blogs, content farms, forums, social media (e.g., Facebook, Reddit, Twitter), IMDb, most wikis (including Wikipedia) and others. Additionally, some sources can be deprecated or blacklisted on Wikipedia.

Given the fact that there are more than 1.5 billion websites on the World Wide Web [3], it is a challenging task to assess the reliability of all of them. Additionally, the reliability is a subjective concept related to information quality [4–6] and each source can be differently assessed depending on topic and language community of Wikipedia. It should also be taken into account that reputation of the newspaper or website can change over time and periodic re-assessment may be necessary.

According to the English Wikipedia content guideline [2]: "in general, the more people engaged in checking facts, analyzing legal issues and scrutinizing the writing, the more reliable the publication." Related work described in Section 2 showed, that there is a field for improving approaches related to assessment of the sources based on publicly available data of Wikipedia using different measures of Wikipedia articles. Therefore, we decided to extract measures related to the demand for information and quality of articles and to use them to build 10 models for assessment of popularity and reliability of the source in different language versions in various periods. The simplest model was based on frequency of occurrence which is commonly used in other related works [7–10]. Other nine novel models used various combinations of measures related to quality and popularity of Wikipedia articles. The models were described in Section 3.

In order to extract sources from references of Wikipedia articles in different languages, we designed and implemented own algorithms in Python. In Section 4 we described basic and complex extraction methods of the references in Wikipedia articles. Based on extracted data from references in each Wikipedia article we added different measures related to popularity and quality of Wikipedia articles (such as pageviews, number of references, article length, number of authors) to assess sources. Based on the results we built rankings of the most popular and reliable sources in different languages editions of Wikipedia. Additionally, we compare positions of selected sources in reliability ranking in different language versions of Wikipedia in Section 5. We also assessed the similarity of the rankings of the most reliable sources obtained by different models in Section 6.

We also designed own algorithms in leveraging data from semantic databases (Wikidata and DBpedia) to extract additional metadata about the sources, conduct their unification and classification to find the most reliable in the specific domains. In Section 7 we showed results of analysis sources based on some parameters from citation templates (such as "publisher" and "journal") and separately we showed the analysis the topics of sources based on semantic databases.

Using different periods we compared the result of popularity and reliability assessment of the sources in Section 8. Comparing the obtained results we were able to find growth leaders described in Section 9. We also presented the assessment of effectiveness of different models in Section 10.1. Additionally we provided information about limitation of the study in Section 10.2.

2. Recent Work

Due to the fact that source reliability is important in terms of quality assessment of Wikipedia articles, there is a wide range of works covering the field of references analysis of this encyclopedia.

Part of studies used reference counts in the models for automatic quality assessment of the Wikipedia articles. One of the first works in this direction used reference count as structural feature to predict the quality of Wikipedia articles [11,12]. Based on the references users can assess the trustworthiness of Wikipedia articles, therefore we consider the source of information as an important factor [13].

Often references contain an external link to the source page (URL), where cited information is placed. Therefore, including in models the number of the external links in Wikipedia articles can also help to assess information quality [14,15].

In addition to the analysis of quantity, there are studies analyzing the qualitative characteristics and metadata related to references. One of the works used special identifiers (such as DOI, ISBN) to unify the references and find the similarity of sources between language versions of Wikipedia [8]. Another recent study analyzed engagement with citations in Wikipedia articles and found that references are consulted more commonly when readers cannot find enough information in selected Wikipedia article [16]. There are also works, which showed that a lot of citations in Wikipedia articles refer to scientific publications [8,17], especially if they are open-access [18], wherein Wikipedia authors prefer to put recently published journal articles as a source [10]. Thus, Wikipedia is especially valuable due to the potential of direct linking to other primary sources. Another popular source of the information in Wikipedia is news website and there is a method for automatic suggestion of the news sources for the selected statements in articles [19].

Reference analysis can be important for quality assessment of Wikipedia articles. At the same time, articles with higher quality must have more proven and reliable sources. Therefore, in order to assess the reliability of specific source, we can analyze Wikipedia articles, in which related references are placed.

Relevance of article length and number of references for quality assessment of Wikipedia content was supported by many publications [15,20–26]. Particularly interesting is the combination of these indicators (e.g., references and articles length ratio) as it can be more actionable in quality prediction than each of them separately [27].

Information quality of Wikipedia depends also on authors who contributed to the article. Often articles with the high quality are jointly created by a large number of different Wikipedia users [28,29]. Therefore, we can use the number of unique authors as one of the measures of quality of Wikipedia articles [26,30,31]. Additionally, we can take into the account information about experience of Wikipedians [32].

One of the recent studies showed that after loading a page, 0.2% of the time the reader clicks on an external reference, 0.6% on an external link and 0.8% hovers over a reference [9]. Therefore, popularity can play an important role not only for quality estimation of information in specific language version of Wikipedia [33] but also for checking reliability of the sources in it. Larger number of readers of a Wikipedia article may allow for more rapid changes in incorrect or outdated information [26]. Popularity of an article can be measured based on the number of visits [34].

Taking into account different studies related to reference analysis and quality assessment of Wikipedia articles, we created 10 models for source assessment. Unlike other studies we used more complex methods of extraction of references and included more language versions of Wikipedia. Additionally, we used semantic layer to identify source type and metadata to create ranking of the sources in specific domains. We also took into account different periods to compare the reliability indicators of the source in various months and to find the growth leaders. Moreover, models were used to assess references based on publicly available data (Wikimedia Downloads [35]), so anybody can use our models for different purposes.

3. Popularity and Reliability Models of the Wikipedia Sources

In this Section we describe ten models related to popularity and reliability of the sources. In most cases source means domain (or subdomain) of the URL in references. Models are identified with abbreviations:

1. **F** model—based on frequency (F) of source usage.
2. **P** model—based on cumulative pageviews (P) of the article in which source appears.
3. **PR** model—based on cumulative pageviews (P) of the article in which source appears divided by number of the references (R) in this article.
4. **PL** model—based on cumulative pageviews (P) of the article in which source appears divided by article length (L).
5. **Pm** model—based on daily pageviews median (Pm) of the article in which source appears.
6. **PmR** model—based on daily pageviews median (Pm) of the article in which source appears divided by number of the references (R) in this article.
7. **PmL** model—based on daily pageviews median (Pm) of the article in which source appears divided by article length (L).
8. **A** model—based on number of authors (A) of the article in which source appears.
9. **AR** model—based on number of authors (A) of the article in which source appears divided by number of the references (R) in this article.
10. **AL** model—based on number of authors (A) of the article in which source appears divided by article length (L).

Frequency of source usage in **F** model means how many references contain the analyzed domain in URL. This method was commonly used in related works [7–10]. Here we take into account a total number of appearances of such reference, i.e., if the same source is cited 3 times, we count the frequency as 3. Equation (1) shows the calculation for **F** model.

$$F(s) = \sum_{i=1}^{n} C_s(i), \tag{1}$$

where s is the source, n is a number of the considered Wikipedia articles, $C_s(i)$ is a number of references using source s (e.q. domain in URL) in article i.

Pageviews, i.e., number of times a Wikipedia article was displayed, is correlated with its quality [33]. We can expect that articles read by many people are more likely to have verified and reliable sources of information. The more people read the article the more people can notice inappropriate source and the faster one of the readers decides to make changes.

P model includes additionally to the frequency of source also cumulative pageviews of the article in which this source appears. Therefore, the source that was mentioned in a reference in a popular article can have bigger value then source that was mentioned even in several less popular articles. Equation (2) presents the calculation of measure using **P** model.

$$P(s) = \sum_{i=1}^{n} C_s(i) \cdot V(i), \tag{2}$$

where s is the source, n is a number of the considered Wikipedia articles, $C_s(i)$ is a number of references using source s (e.q. domain in URL) in article i, $V(i)$ is cumulative pageviews value of article i.

PR model uses cumulative pageviews divided by the total number of the references in a considered article. Unlike the previous model here we take into account visibility of the references using the

analyzed source. We assume that in general the more references in the article, the less visible the specific reference is Equation (3) shows the calculation of measure using **PR** model.

$$PR(s) = \sum_{i=1}^{n} \frac{V(i)}{C(i)} \cdot C_s(i), \tag{3}$$

where s is the source, n is a number of the considered Wikipedia articles, $C(i)$ is total number of the references in article i, $C_s(i)$ is a number of the references using source s (e.q. domain in URL) in article i, $V(i)$ is cumulative pageviews value of article i.

Another important aspect of the visibility of each reference is the length of the entire article. Therefore, we provide additional **PL** model that operates on the principles described in Equation (4).

$$PL(s) = \sum_{i=1}^{n} \frac{V(i)}{T(i)} \cdot C_s(i), \tag{4}$$

where s is the source, n is a number of the considered Wikipedia articles, $T(i)$ is the length of source code (wiki text) of article i, $C_s(i)$ is a number of references using source s (e.q. domain in URL) in article i, $V(i)$ is cumulative pageviews value of article i.

Popularity of an article can be measured in different ways. As it was proposed in [26] we decided to measure pageviews also as daily pageviews median (Pm) of individual articles. Thereby we provided additional models **Pm**, **PmR**, **PmL** that are modified versions of models **P**, **PR**, **PL**, respectively. The modification consists in replacement of cumulative pageviews with daily pageviews median.

As the pageviews value of article is more related to readers, we also propose a measure addressing the popularity among authors, i.e., number of users who decided to add content or make changes in the article. Given the assumptions of previous models we propose analogous models related to authors: models **A**, **AR**, **AL** are described in Equations (5)–(7), respectively.

$$A(s) = \sum_{i=1}^{n} C_s(i) \cdot E(i), \tag{5}$$

where s is the source, n is a number of the considered Wikipedia articles, $C_s(i)$ is a number of references using source s (e.q. domain in URL) in article i, $E(i)$ is total number of authors of article i.

$$AR(s) = \sum_{i=1}^{n} \frac{E(i)}{C(i)} \cdot C_s(i), \tag{6}$$

where s is the source, n is a number of the considered Wikipedia articles, $C(i)$ is total number of the references in article i, $C_s(i)$ is a number of references using source s (e.q. domain in URL) in article i, $E(i)$ is total number of authors of article i.

$$AL(s) = \sum_{i=1}^{n} \frac{E(i)}{T(i)} \cdot C_s(i), \tag{7}$$

where s is the source, n is a number of the considered Wikipedia articles, $T(i)$ is the length of source code (wiki text) of article i, $C_s(i)$ is a number of references using source s (e.q. domain in URL) in article i, $E(i)$ is total number of authors of article i.

It is important to note that for pageviews measures connected with sources extracted in the end of the assessed period we use data for the whole period (month). For example, if references were extracted based on dumps as of 1 March 2020, then we considered pageviews of the articles for the whole February 2020.

4. Extraction of Wikipedia References

Wikimedia Foundation back-ups each language version of Wikipedia at least once a month and stores it on a dedicated server as "Database backup dumps". Each file contains different data related to Wikipedia articles. Some of them contain source codes of the Wikipedia pages in wiki markup, some of them describe individual elements of articles: headers, category links, images, external or internal links, page information and others. There are even files that contain the whole edit history of each Wikipedia page.

Variety of dump files gives possibility to extract necessary data in different ways. Some of them allow to get results in a relatively short time using simple parser. However, other important information may be missing in such files. Therefore, in this section we describe two methods of extracting the data about references in Wikipedia.

4.1. Basic Extraction

References have often links to different external sources (websites). For each language version of Wikipedia we used dump file with external URL link records in order to extract the URLs from rendered versions of Wikipedia article. For instance, for English Wikipedia we used dump file from March 2020-"enwiki-20200301-externallinks.sql.gz". This file contains data about external links placed in all pages in selected language version of Wikipedia. Therefore, we took into account only links placed in article namespace (ns0). We extracted over 280 million external links from 55 considered language versions of Wikipedia. Table 1 shows the extraction statistics based on dumps from March 2020: total number of articles, number of articles with a certain number of external links (URLs), total and unique number of external links in different language versions of Wikipedia.

Analysis of the external links showed that the largest share of articles with at least one link is placed in Swedish Wikipedia—96%. English Wikipedia has slightly less value of this indicator—about 91% articles with at least 1 external link. However, English Wikipedia has the largest share of articles with at least 100 external links—1% of all articles in this language. The biggest total number of external links per 1 article has Catalan (12.7), English (11.5) and Russian (10.1) Wikipedia.

Based on the extraction of external links, we can find which of the domains (or subdomains) are often used in Wikipedia articles. Figure 1 shows the most popular domains (and subdomains) in over 280 million external links from 55 language versions of Wikipedia.

Figure 1. The most popular domains in over 280 million external links from 55 language versions of Wikipedia. Source: own calculations based on Wikimedia Dumps as of March 2020 using basic extraction method. The most popular domains in external links in other language versions are available on the web page: http://data.lewoniewski.info/sources/basic.

It is important to note that despite the fact that imdb.com (Internet Movie Database) included in the list of sites which are generally unacceptable in English Wikipedia [2], this resource is on the 2nd planes in the list of the most commonly used websites in Wikipedia articles. The top 10 of the most commonly used websites also contains: web.archive.org (Wayback Machine), viaf.org (Virtual International Authority File), int.soccerway.com (Soccerway-website on football), tvbythenumbers.zap2it.com (TV by the Numbers), animaldiversity.org (Animal Diversity Web), deadline.com (Deadline Hollywood),

variety.com (Variety-american weekly entertainment magazine), webcitation.org (WebCite-on-demand archiving service), officialcharts.com (The Official UK Charts Company).

Table 1. Total number of articles, number of articles with a certain number of external links (URLs), total and unique number of external links in different language versions of Wikipedia. Source: own calculations based on Wikimedia dumps in March 2020 using complex extraction of references.

Language	Number of Articles				Number of URLs	
	All	with >=1 URL	>= 10 URLs	>=100 URLs	All	Unique
ar (Arabic)	1,031,740	917,809	305,118	4369	9,443,788	7,599,390
az (Azerbaijani)	156,442	109,743	20,299	237	674,212	512,465
be (Belarusian)	185,753	150,116	21,067	299	1,142,005	958,165
bg (Bulgarian)	260,081	211,031	27,806	185	1,174,324	1,030,715
ca (Catalan)	638,664	600,711	336,302	1770	8,111,104	7,124,746
cs (Czech)	447,120	377,647	69,821	1220	2,769,415	2,438,870
da (Danish)	257,321	211,415	51,689	488	1,711,677	1,605,379
de (German)	2,403,683	1,990,310	528,524	7849	17,646,882	15,632,584
el (Greek)	174,589	151,008	43,664	891	1,479,933	1,254,224
en (English)	6,029,201	5,500,527	1,963,703	60,384	69,554,575	56,030,670
eo (Esperanto)	275,674	223,652	21,028	85	1,016,902	928,935
es (Spanish)	1,528,811	1,395,107	484,650	5521	13,935,332	11,872,312
et (Estonian)	206,430	136,651	8,344	146	526,292	466,916
eu (Basque)	349,176	331,836	97,469	104	2,692,639	2,177,612
fa (Persian)	712,216	656,161	52,779	1030	2,779,293	2,232,907
fi (Finnish)	479,830	405,372	61,387	545	2,446,538	1,889,702
fr (French)	2,185,885	1,830,876	593,874	7327	17,918,673	15,313,234
gl (Galician)	161,860	127,395	52,159	595	1,483,541	1,315,467
he (Hebrew)	261,209	213,989	76,274	347	2,152,942	1,987,360
hi (Hindi)	140,327	97,706	10,102	370	563,963	379,306
hr (Croatian)	198,670	137,949	10,796	155	587,017	449,783
hu (Hungarian)	465,509	411,072	97,289	1179	3,231,880	2,796,234
hy (Armenian)	264,676	219,045	50,681	1218	2,073,940	1,534,220
id (Indonesian)	524,100	409,937	53,085	1267	2,496,158	2,158,397
it (Italian)	1,586,855	1,374,018	403,171	3194	11,889,377	10,141,992
ja (Japanese)	1,192,596	890,138	205,264	4210	7,449,642	6,309,830
ka (Georgian)	135,333	102,910	10,508	239	533,019	420,322
kk (Kazakh)	230,376	137,333	6,536	54	736,786	591,481
ko (Korean)	486,067	318,190	63,425	1110	2,197,777	1,990,960
la (Latin)	132,258	106,887	3,592	22	347,131	287,532
lt (Lithuanian)	196,606	136,982	4,238	27	390,006	331,424
ms (Malay)	335,222	191,206	18,288	431	868,166	716,712
nl (Dutch)	1,999,092	1,626,602	31,700	1460	4,303,813	3,295,204
nn (Norwegian (Nynorsk))	151,857	126,229	16,642	73	624,568	561,283
no (Norwegian)	529,426	466,557	132,817	672	3,812,791	3,410,905
pl (Polish)	1,387,164	1,177,588	159,956	2334	6,962,407	5,673,526
pt (Portuguese)	1,022,524	925,771	186,889	4454	7,836,416	6,583,420
ro (Romanian)	404,748	352,338	80,111	970	2,742,321	2,375,095
ru (Russian)	1,602,761	1,333,264	527,323	8184	16,116,795	12,370,583
sh (Serbo-Croatian)	451,298	383,945	223,652	292	4,464,569	1,118,996
simple (Simple English)	155,887	103,886	10,990	264	548,488	480,654
sk (Slovak)	232,551	176,188	10,893	268	823,474	681,781
sl (Slovenian)	167,119	135,614	21,910	219	786,235	710,113
sr (Serbian)	630,870	552,584	53,185	761	3,502,213	1,959,054
sv (Swedish)	3,740,411	3,590,906	798,561	2356	21,372,068	11,686,205
ta (Tamil)	132,424	105,186	10,658	228	569,482	401,066
th (Thai)	135,627	93,945	16,965	726	758,451	667,308
tr (Turkish)	343,216	257,976	40,305	1306	1,762,805	1,495,178
uk (Ukrainian)	994,030	859,711	185,470	2476	6,973,455	5,195,088
ur (Urdu)	154,282	120,189	5229	191	403,727	354,010
uz (Uzbek)	133,774	92,369	964	27	299,080	265,877
vi (Vietnamese)	1,241,487	1,178,177	46,835	1580	3,604,033	2,846,271
vo (Volapük)	124,189	93,924	9	-	104,201	103,660
zh (Chinese)	1,099,744	862,260	175,496	4873	6,757,646	5,779,801
zh-min-nan (Min Nan)	267,615	192,933	519	1	353,098	274,056

Obtained results can be used for further analysis. However, basic extraction method next to its its relative simplicity, have some disadvantages. For example, we can extract all external links from article using basic extraction method but we will miss information about placement of each link in article (e.q. if it was placed in reference). Another problem is excluding not relevant links such as archived copy of the source (when the original copy in presented and available), links generated automatically if the source has special identifiers or templates, links to other pages of Wikimedia projects (often they show additional information about the article but not the source of information) and others. Therefore, we decided to conduct a more complex extraction based on source code of each Wikipedia article. This method is described in the next subsection.

4.2. Complex Extraction

Using Wikipedia dumps from March 2020, we have extracted all references from over 40 million articles in 55 language editions that have at least 100,000 articles and at least 5 article depth index in recent years as it was proposed in [26]. Complex extraction was based on source code of the articles. Therefore, we used other dump file (comparing to basic extraction)-for example dump file as of March 2020 for English Wikipedia that we used is "enwiki-20200301-pages-articles.xml.bz2".

In wiki-code references are usually placed between special tags <*ref*>. . . </*ref*>. Each reference can be named by adding "name" parameter to this tag: <*ref name="..."*>...</*ref*>. After such reference was defined in the articles, it can be placed elsewhere in this article using only <*ref name="..."* />. This is how we can use the same reference several times using default wiki markup. However, there are other possibilities to do so. Depending on language version of Wikipedia we can also use special templates with specific names and set of parameters. It is not even mandatory that some of them must be placed under <*ref*>...</*ref*> tag.

In general, we can divide references into two groups: with special template and without it. In the case of references without special template they usually have URL of source and some optional description (e.g., title). References with special templates can have different data describing the source. Here in separate fields one can add information about author(s), title, URL, format, access date, publisher and others. The set of possible parameters with predefined names depends on language version and type of templates, which can describe book, journal, web source, news, conference and others. Figure 2 shows the most commonly used templates in <*ref*> tags in English. Among the most commonly used templates in this Wikipedia language versions are: 'Cite web', 'Cite news', 'Cite book', 'Cite journal', National Heritage List for England ('NHLE'), 'Citation', 'Webarchive', 'ISBN', 'In lang', 'Dead link', Harvard citation no brackets ('Harvnb'), 'Cite magazine'. In order to extract information about sources we created own algorithms that take into account different names of reference templates and parameters in each language version of Wikipedia. The most commonly used parameters in this language version are: title, url, accessdate, date, publisher, last, first, work, website and access-date.

Figure 2. The most popular templates used in references in English Wikipedia. Source: own calculations based on Wikimedia Dumps as of March 2020. The most popular templates in other language versions are presented on the web page: http://data.lewoniewski.info/sources/templates.

It is important to note that the presence of some references cannot be identified directly based on the source (wiki) code of the articles. Sometimes infoboxes or other templates in the Wikipedia

article can put additional references to the rendered version of article. Figure 3 shows such situation on example of table with references in the Wikipedia article "2019–2020 coronavirus pandemic" that was added using template "2019–2020 coronavirus pandemic data". In our approach we include such references in the analysis when such templates appear in the Wikipedia articles.

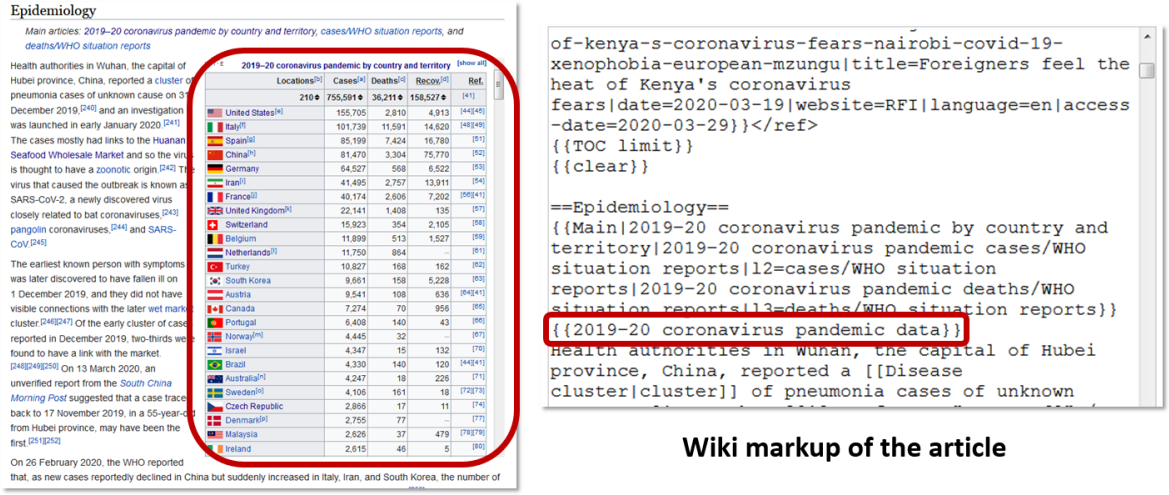

Figure 3. Table with references in the Wikipedia article 2019–2020 coronavirus pandemic that was added using template 2019–2020 coronavirus pandemic data. Source [36].

Some of the most popular templates allows to add identifiers to the source such as DOI, JSTOR, PMC, PMID, arXiv, ISBN, ISSN, OCLC and others. Some references can include special templates related to identifiers such DOI, ISBN, ISSN can be described as separate templates. For example, value for "doi" parameter can be written as "doi|...". Moreover, some of the templates allow to insert several identifiers for one reference-templates for ISBN, ISSN identifiers allows to put two or more values-for example we can put in code "ISBN|...|..." or "ISSN|...|...|...". Table 2 shows the extraction statistics of the references with DOI, ISBN, ISSN, PMID, PMC identifiers. Table 3 shows the extraction statistics of the references with arXiv, Bibcode, JSTOR, LCCN, OCLC identifiers.

Table 2. Total and unique number of references with special identifiers: DOI, ISBN, ISSN, PMID, PMC. Source: own calculations based on Wikimedia dumps as of March 2020 using complex extraction of references.

Language	DOI		ISBN		ISSN		PMID		PMC	
	All	Unique	All	Unique	All	Unique	All	Unique	All	Unique
ar (Arabic)	130,246	87,431	169,583	78,655	24,038	7718	83,228	58,097	18,374	12,793
az (Azerbaijani)	2290	1320	23,303	8823	903	260	540	383	128	107
be (Belarusian)	2494	1568	48,314	6426	1049	288	1120	735	165	111
bg (Bulgarian)	8431	5823	53,738	14,536	1345	503	6024	3745	989	700
ca (Catalan)	49,819	34,451	226,677	76,508	25,939	6878	27,678	21,796	7829	6343
cs (Czech)	26,413	15,891	177,252	33,402	28,785	4659	12,271	7795	1318	925
da (Danish)	7440	4619	32,223	13,041	1522	540	4879	2859	892	556
de (German)	158,399	82,168	890,727	199,949	77,065	13,250	18,893	12,821	14,660	9284
el (Greek)	22,803	14,416	66,982	27,292	4751	1541	12,325	7758	2509	1647
en (English)	2,130,154	919,480	4,374,241	848,284	550,834	39,487	993,092	477,883	346,934	156,941
eo (Esperanto)	4806	3177	18,128	9464	687	332	1922	1249	565	340
es (Spanish)	136,761	77,866	653,902	168,306	97,688	14,201	65,499	39,328	14,867	8457
et (Estonian)	7481	3269	16,650	5148	534	171	4809	2063	1134	509
eu (Basque)	7136	5115	17,159	9413	6811	2202	1511	1190	5938	3633
fa (Persian)	27,025	18,631	45,849	20,908	5212	2146	17,180	11,371	4499	3002
fi (Finnish)	10,151	5394	177,952	25,085	7991	1954	5182	2936	370	276
fr (French)	164,244	73,634	954,903	201,227	125,842	15,592	40,177	26,035	4294	2307
gl (Galician)	45,231	30,385	68,558	21,678	6351	1936	34,907	24,796	10,275	7092
he (Hebrew)	8611	7751	12,953	9661	1599	418	3883	3632	590	546

Table 2. *Cont.*

Language	DOI		ISBN		ISSN		PMID		PMC	
	All	Unique	All	Unique	All	Unique	All	Unique	All	Unique
hi (Hindi)	10,907	6988	27,212	12,300	1011	439	9168	6213	1442	1019
hr (Croatian)	5570	3318	20,663	8053	903	292	4872	2726	726	471
hu (Hungarian)	21,998	14,473	65,548	20,596	4245	1528	12,139	8154	2090	1466
hy (Armenian)	36,918	22,192	59,679	25,857	6585	2348	32,282	19,084	7628	4568
id (Indonesian)	36,819	21,546	139,969	46,121	9416	2455	16,645	10,743	3957	2709
it (Italian)	90,207	51,150	423,668	95,214	13,880	3422	45,394	32,052	7030	4642
ja (Japanese)	119,914	58,187	573,918	92,362	29,754	5567	42,793	27,130	11,225	6598
ka (Georgian)	3592	2541	15,425	6310	1228	339	1394	1084	297	238
kk (Kazakh)	375	312	55,956	1346	67	39	210	178	78	63
ko (Korean)	40,529	20,761	62,384	23,804	6355	1640	14,499	9374	3888	2453
la (Latin)	700	521	2870	1860	49	36	294	245	74	55
lt (Lithuanian)	1456	1083	10,851	3597	316	138	940	655	187	144
ms (Malay)	11,423	7681	30,838	14,583	1931	694	5678	3931	1416	945
nl (Dutch)	12,669	8538	45,588	16,296	1782	821	7496	5339	1470	1036
nn (Nynorsk)	3412	1789	19,903	6649	611	180	770	479	165	108
no (Norwegian)	11,868	6521	69,350	25,932	7054	1391	4924	3169	734	510
pl (Polish)	131,704	42,238	519,934	62,974	74,490	9111	49,274	28,002	6944	3716
pt (Portuguese)	84,664	45,575	263,774	81,583	34,965	7029	33,826	20,825	7123	4466
ro (Romanian)	18,715	11,592	62,057	22,448	3114	1048	10,780	6488	2260	1507
ru (Russian)	133,388	63,725	639,602	131,769	68,765	11,259	37,716	23,119	7328	4422
sh (Serbo-Croatian)	53,965	12,922	44,875	11,669	3374	657	29,012	21,927	3227	2225
simple (Simple Engl.)	7730	5337	26,299	13,472	2103	612	4265	2953	908	668
sk (Slovak)	3166	2238	40,302	8914	7047	1131	735	569	127	106
sl (Slovenian)	12,819	7943	44,213	12,273	1579	674	9541	5994	1689	1104
sr (Serbian)	67,079	22,115	94,352	29,576	6807	2007	35,541	26,308	5011	3336
sv (Swedish)	863,337	8954	145,177	27,169	11,501	2399	6605	3694	1336	817
ta (Tamil)	19,679	14,006	28,470	15,720	1714	762	11,131	8164	2088	1386
th (Thai)	26,449	16,162	32,288	14,812	2879	937	18,959	11,730	4251	2681
tr (Turkish)	18,681	11,118	54,775	22,021	3491	1027	9360	6021	1739	1202
uk (Ukrainian)	255,144	24,659	122,999	37,644	53,699	3349	55,249	10,143	3224	2216
ur (Urdu)	1481	897	8389	4735	362	138	546	379	157	106
uz (Uzbek)	144	126	870	542	25	19	26	24	11	10
vi (Vietnamese)	71,162	39,745	139,027	44,896	10,740	2773	32,729	21,701	8674	5665
vo (Volapük)	-	-	87	77	-	-	-	-	-	-
zh (Chinese)	109,034	59,455	362,310	92,422	25,637	6349	48,215	29,791	11,077	6883
zh-min-nan (Min Nan)	290	163	618	262	20	11	62	51	20	15

Table 3. Total and unique number of references with special identifiers: arXiv, Bibcode, JSTOR, LCCN, OCLC. Source: own calculations based on Wikimedia dumps as of March 2020 using complex extraction of references.

.

Language	arXiv		Bibcode		JSTOR		LCCN		OCLC	
	All	Unique	All	Unique	All	Unique	All	Unique	All	Unique
ar (Arabic)	8604	3016	21,129	9943	5123	3693	425	287	8370	5091
az (Azerbaijani)	144	72	797	392	474	141	183	75	504	174
be (Belarusian)	253	129	547	318	52	38	4	3	80	41
bg (Bulgarian)	404	309	1395	1089	298	227	104	47	799	344
ca (Catalan)	1735	911	5562	3352	1641	1025	190	101	4018	1543
cs (Czech)	1436	580	3817	1713	207	139	24	17	7425	2516
da (Danish)	161	85	755	541	246	170	97	40	1254	399
de (German)	6430	3318	7586	3591	3789	2060	266	116	4633	2516
el (Greek)	1829	871	5262	2963	970	523	140	57	1477	631
en (English)	154,579	28,727	396,409	117,983	169,419	71,447	22,374	4921	312,755	79,862
eo (Esperanto)	39	20	241	179	356	253	22	21	199	155
es (Spanish)	2914	1653	12,188	7252	6713	3858	671	267	60,597	14,105
et (Estonian)	320	132	1355	597	134	68	9	5	148	76
eu (Basque)	185	51	387	165	53	46	4	3	173	118
fa (Persian)	898	533	3200	2133	679	529	95	42	2005	875
fi (Finnish)	110	89	460	345	164	104	38	28	133	100
fr (French)	11,448	3043	23,513	7147	5345	2755	7653	2616	77,037	23,267
gl (Galician)	831	340	3894	2323	1524	841	678	207	5810	1411
he (Hebrew)	70	68	344	315	245	226	9	9	1599	1219
hi (Hindi)	1063	273	2189	775	222	157	55	33	619	349
hr (Croatian)	166	124	688	521	117	80	14	4	396	162
hu (Hungarian)	357	243	1602	1164	620	461	59	42	1282	481
hy (Armenian)	448	255	3666	1681	762	550	86	38	1905	849
id (Indonesian)	2314	819	7784	3638	2405	1198	489	160	10,039	2717

Table 3. *Cont.*

Language	arXiv		Bibcode		JSTOR		LCCN		OCLC	
	All	Unique	All	Unique	All	Unique	All	Unique	All	Unique
it (Italian)	2846	1291	5860	3610	1916	1138	2419	683	20,114	7209
ja (Japanese)	11,253	3075	33,245	9453	2755	1543	811	234	12,169	3876
ka (Georgian)	425	269	1143	802	157	115	82	52	465	250
kk (Kazakh)	36	20	55	50	17	14	-	-	20	16
ko (Korean)	7621	2565	15,160	5517	1281	837	374	114	1529	623
la (Latin)	4	4	52	45	47	34	6	6	44	30
lt (Lithuanian)	122	79	196	147	47	38	1	1	105	49
ms (Malay)	657	374	2386	1570	528	360	57	43	2337	646
nl (Dutch)	35	28	317	261	163	123	22	5	336	271
nn (Norwegian (Nynorsk))	749	169	1649	570	195	98	19	5	223	123
no (Norwegian)	975	252	3027	1258	392	249	43	26	1547	611
pl (Polish)	2493	863	5414	2215	1261	635	223	79	28,195	7579
pt (Portuguese)	4260	1666	19,602	6013	2844	1806	321	170	11,514	4891
ro (Romanian)	1181	495	4004	2270	681	480	175	85	2008	800
ru (Russian)	12,622	3301	25,754	7756	2358	1288	368	131	4988	1772
sh (Serbo-Croatian)	171	91	1101	720	401	295	37	17	3435	787
simple (Simple English)	544	265	1222	825	227	177	46	26	1246	404
sk (Slovak)	198	131	398	291	24	17	10	3	334	160
sl (Slovenian)	664	187	1473	676	298	261	40	16	504	330
sr (Serbian)	637	415	2982	2072	975	718	129	87	4221	1914
sv (Swedish)	1042	391	3097	1257	311	223	198	22	5105	1629
ta (Tamil)	699	306	2625	1663	547	372	84	43	895	475
th (Thai)	1053	340	2859	1506	492	326	37	26	997	507
tr (Turkish)	2150	769	5282	2395	701	380	107	59	1550	588
uk (Ukrainian)	4327	1754	14,628	5243	943	660	214	89	3011	1450
ur (Urdu)	93	33	208	127	141	104	9	7	385	158
uz (Uzbek)	24	20	93	79	6	5	16	4	14	11
vi (Vietnamese)	7798	2644	18,881	7895	2568	1462	342	208	6779	1895
vo (Volapük)	-	-	-	-	-	-	-	-	-	-
zh (Chinese)	11,497	3496	27,199	10,459	2819	1623	589	255	12,360	3482
zh-min-nan (Min Nan)	1	1	82	43	9	7	-	-	9	9

Special identifiers can determine similarity between the references even though they have different parameters in description (e.g., titles in another languages). Unification of these references can be done based on identifiers. For example, if a reference has DOI number "10.3390/computers8030060", we give it URL "https://doi.org/10.3390/computers8030060". More detailed information about identifiers which we used to unifying the references is shown in Table 4.

Table 4. Identifiers that were used for URL unification of references.

Identifier	Description	URL
arXiv	arXiv repository identifier	https://arxiv.org/abs/...
Bibcode	Compact identifier used by several astronomical data systems	https://adsabs.harvard.edu/abs/...
DOI	Digital object identifier	https://doi.org/...
ISBN	International Standard Book Number	https://books.google.com/books?vid=ISBN...
ISSN	International Standard Serial Number	https://worldcat.org/ISSN/...
JSTOR	Journal Storage number	https://jstor.org/stable/...
LCCN	Library of Congress Control Number	https://lccn.loc.gov/
PMC	PubMed Central	https://ncbi.nlm.nih.gov/pmc/articles/PMC...
PMID	PubMed	https://ncbi.nlm.nih.gov/pubmed/...
OCLC	WorldCat's Online Computer Library Center	https://worldcat.org/oclc/...

One of the advantages of the complex method of extraction (comparing to basic one, which was described in previous subsection) is ability to distinguish between types of source URLs: actual link to the page and archived copy. For linking to web archiving services such as the Wayback Machine, WebCite and other web archiving services special template "Webarchive" can be used. In most cases the template needs only two arguments, the archive url and date. This template is used in different languages and sometimes has different names. Additionally, in a single language this template can

be called using other names, which are redirects to original one. For example in English Wikipedia alternative names of this templates can be used: "Weybackdate", "IAWM", "Webcitation", "Wayback", "Archive url", "Web archive" and others. Using information from those templates we found the most frequent domains of web archiving services in references.

It is important to note that depending on language version of Wikipedia template about archived URL addresses can have own set of parameters and own way to generate final URL address of the link to the source. For example, in the English Wikipedia template Webarchive has parameter url which must contain full URL address from web archiving service. At the same time related template Webarchiv in German Wikipedia has also other ways to define a link to archived source-one can provide URL of the original source page (that was created before it was archived) using url parameter and (or) additionally use parameters depending on the archive service: "wayback", "archive-is", "webciteID" and others. In this case, to extract the full URL address of the archived web page, we need to know how inserted value of each parameter affects the final link for the reader of the Wikipedia article in each language version.

In the extraction we also took into account short citation from "Harvard citation" family of templates which uses parenthetical referencing. These templates are generally used as in-line citations that link to the full citation (with the full meta data of the source). This enables a specific reference to be cited multiple times having some additional specification (such as a page number) with other details (comments). We included in the analysis following templates: "Harvnb" (Harvard citation), "harvnb" (Harvard citation no brackets), "Harvtxt" (Harvard citation text), "Harvcol", "Harvcolnb", "Sfn" (Shortened footnote template) and others. Depending on language version of Wikipedia, each template can have another corresponding name and additional synonymous names. For example in English Wikipedia, "Harvard citation", "Harv" and "Harvsp" mean the same template (with the same rules), while corresponding template in French has such names as "Référence Harvard", "Harvard" and also "Harv".

Taking into account unification of URLs based on special identifiers, excluding URLs of archived copies of the sources and including special templates outside *<ref>* tags, we counted the number of all and unique references in each considered language version. Table 5 presents total number of articles, number of articles with at least 1 reference, at least 10 references, at least 100 references and number of total and unique number of references in each considered language version of Wikipedia.

Table 5. Total number of articles, number of articles with at least 1 reference, at least 10 references, at least 100 references and number of total and unique number of references in each considered language version of Wikipedia. Source: own calculation based on Wikimedia dumps as of March 2020 using complex extraction of references.

Language	Number of Articles				Number of References	
	All	with >= 1 ref.	with >= 10 refs.	with >= 100 refs.	All	Unique
ar (Arabic)	1,031,740	817,485	58,303	2588	3,598,691	2,138,127
az (Azerbaijani)	156,442	77,213	6476	440	430,655	210,186
be (Belarusian)	185,753	90,427	5897	269	352,275	163,649
bg (Bulgarian)	260,081	152,632	13,099	330	702,747	397,568
ca (Catalan)	638,664	421,096	55,870	1443	2,676,870	1,334,484
cs (Czech)	447,120	229,700	45,793	1267	1,762,136	911,167
da (Danish)	257,321	99,188	13,157	615	614,575	395,741
de (German)	2,403,683	1,350,469	276,204	6214	10,343,100	6,150,128
el (Greek)	174,589	100,645	24,080	1000	971,438	589,234
en (English)	6,029,201	4,738,526	1,363,475	67,179	58,914,062	28,973,680
eo (Esperanto)	275,674	54,839	4091	149	230,042	152,878
es (Spanish)	1,528,811	1,078,622	233,774	54,963	14,428,514	4,495,443
et (Estonian)	206,430	90,628	11,709	392	568,263	258,665
eu (Basque)	349,176	157,679	4045	116	563,102	172,629
fa (Persian)	712,216	383,131	23,183	1111	1,393,976	840,009
fi (Finnish)	479,830	340,425	65,714	1464	2,514,637	1,198,430
fr (French)	2,185,885	1,290,227	314,893	12,303	12,407,709	6,477,543
gl (Galician)	161,860	73,040	12,476	560	560,381	297,875

Table 5. *Cont.*

Language	Number of Articles				Number of References	
	All	with >= 1 ref.	with >= 10 refs.	with >= 100 refs.	All	Unique
he (Hebrew)	261,209	126,063	24,712	360	895,644	777,279
hi (Hindi)	140,327	55,173	6354	403	331,919	203,538
hr (Croatian)	198,670	100,749	8457	313	463,336	246,391
hu (Hungarian)	465,509	174,547	41,585	1292	1,433,477	817,002
hy (Armenian)	264,676	182,065	19,299	937	984,768	528,465
id (Indonesian)	524,100	226,673	33,691	1542	1,525,411	845,109
it (Italian)	1,586,855	698,996	143,034	5406	5,895,516	3,273,847
ja (Japanese)	1,192,596	694,366	206,822	10,229	8,701,385	4,485,637
ka (Georgian)	135,333	46,308	4945	290	265,153	160,032
kk (Kazakh)	230,376	144,401	1011	48	274,529	52,503
ko (Korean)	486,067	170,646	24,467	1006	1,136,561	725,725
la (Latin)	132,258	45,476	1563	27	128,992	66,105
lt (Lithuanian)	196,606	68,043	3221	48	212,662	143,521
ms (Malay)	335,222	76,845	10,534	469	487,718	311,772
nl (Dutch)	1,999,092	956,918	27,768	619	2,082,368	1,198,126
nn (Norwegian (Nynorsk))	151,857	44,191	4588	126	220,340	125,740
no (Norwegian)	529,426	253,183	23,932	953	1,243,303	691,525
pl (Polish)	1,387,164	802,519	160,599	4168	6,035,345	2,467,049
pt (Portuguese)	1,022,524	728,219	103,344	5480	4,944,321	2,710,271
ro (Romanian)	404,748	232,248	32,527	1295	1,481,560	625,841
ru (Russian)	1,602,761	978,601	219,135	8083	8,857,326	4,610,614
sh (Serbo-Croatian)	451,298	338,340	15,451	393	1,322,980	214,925
simple (Simple English)	155,887	81,691	8799	311	431,401	274,407
sk (Slovak)	232,551	89,345	8027	228	411,430	224,896
sl (Slovenian)	167,119	64,210	7717	343	367,513	197,430
sr (Serbian)	630,870	494,946	18,870	816	2,789,499	487,172
sv (Swedish)	3,740,411	3,123,685	135,228	33,497	20,053,493	4,207,630
ta (Tamil)	132,424	91,023	8981	280	490,602	255,568
th (Thai)	135,627	69,954	12,634	642	581,563	362,812
tr (Turkish)	343,216	163,287	22,472	1091	1,121,121	690,471
uk (Ukrainian)	994,030	579,407	81,908	1681	3,894,437	1,417,597
ur (Urdu)	154,282	114,666	3214	185	259,328	194,444
uz (Uzbek)	133,774	25,082	585	31	55,673	23,288
vi (Vietnamese)	1,241,487	1,053,266	41,640	1879	2,747,781	1,602,977
vo (Volapük)	124,189	655	9	-	1525	1374
zh (Chinese)	1,099,744	630,774	112,953	5287	5,009,984	2,740,728
zh-min-nan (Min Nan)	267,615	40,194	161	2	61,896	4898

Analysis of the numbers of the references extracted by complex extraction showed other statistics comparing to basic extraction of the external links described in Section 4.1. The largest share of the article with at least one references has Vietnamese Wikipedia-84.8%. Swedish, Arabic, English and Serbian Wikipedia has 83.5%, 79.2%, 78.2% and 78.1% share of such articles, respectively. If we consider only articles with at least 100 references, then the largest share of such articles will have Spanish Wikipedia-3.5%. English, Swedish and Japanese Wikipedia has 1.1%, 0.9% and 0.8% share of such articles, respectively. However, the largest total number of the references per number of articles has English Wikipedia—9.6 references. Relatively large number of references per article has also Spanish (9.2) and Japanese (7.1) Wikipedia.

The largest number of the references with DOI identifier has English Wikipedia (over 2 million) at the same time has the largest number of average number of references with DOI per article—34.3%. However, the largest share of the references with DOI among all references has Galician (8.4%) and Ukrainian (6.6%) Wikipedia.

The largest number of the references with ISBN identifier has English Wikipedia (over 3.5 million) at the same time has the largest number of average number of references with ISBN per article-34.3%. However, the largest share of the references with ISBN among all references has Kazakh (20.3%) and Belarusian (13.1%) Wikipedia.

Based on the extraction of URLs from the obtained references, we can find which of the domains (or subdomains) are often used in Wikipedia articles. Figure 4 shows the most

popular domains (and subdomains) in over 200 million references of Wikipedia articles in 55 language versions. Comparing results with basic extraction (see Section 4.1) we got some changes in the top 10 of the most commonly used sources in references: deadline.com (Deadline Hollywood), tvbythenumbers.zap2it.com (TV by the Numbers), variety.com (Variety-american weekly entertainment magazine), imdb.com (Internet Movie Database), newspapers.com (historic newspaper archive), int.soccerway.com (Soccerway-website on football), web.archive.org (Wayback Machine), oricon.co.jp (Oricon Charts), officialcharts.com (The Official UK Charts Company), gamespot.com (GameSpot-video game website).

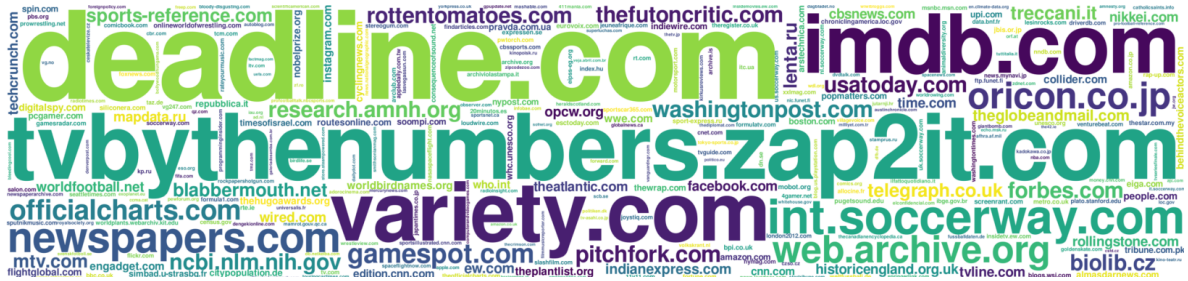

Figure 4. The most popular domains in URL of references of Wikipedia articles in 55 language versions. Source: own calculations based on Wikimedia Dumps as of March 2020 using complex extraction method.

5. Assessment of Sources

To assess the references based on prooped models apart from extraction of the source we also extracted data related to pageviews, lenght of the articles and number of the authors. We used different dumps files that are available on "Wikimedia Downloads" [35].

Based on complex extraction method we measure popularity and reliability of the sources in references. Due to limitation of the size in this paper we often used **F** or **PR** model to show various ranking of sources. The exception is situations where we compared 10 proposed models for popularity and reliability assessment of the sources in Wikipedia. Additionally in the tables we limit number of the languages to one of the most developed: Arabic (ar), German (de), English (en), Spanish (es), Persian (fa), French (fr), Italian(it), Japanese(ja), Dutch (nl), Polish (pl), Portuguese (pt), Russian (ru), Swedish (sv), Vietnamese (vi), Chinese (zh). The more extended version of the results are placed on the web page: http:/data.lewoniewski.info/sources/. For example, figures that shows the most popular and reliable sources for each of considered language version of Wikipedia using F-model (http://data.lewoniewski.info/sources/modelf) and PR-model (http://data.lewoniewski.info/sources/modelpr) placed there.

Table A1 shows position in the local rankings of the most popular and reliable sources in one of the most developed language versions of Wikipedia in February 2020 using **PR** model. In this table it is possible to compare rank of the source that has leading position in at least one language version to other languages. For example, "taz.de" (Die Tageszeitung) is on 3rd place in German Wikipedia in February 2020, at the same time this source is on 692nd, 785th and 996th place in French, Persian and Polish Wikipedia respectively in the same period. In French Wikipedia the most reliable source in February 2020 was "irna.ir" (Islamic Republic News Agency), at the same time in English Wikipedia it is on 8072nd place. However this source not mentioned at all in Polish and Swedish Wikipedia. Other example-in Russian Wikipedia the most reliable source in February 2020 "lenta.ru" was on the 1st place, at the same time it is on the 166th, 310th, 325th and 352nd in Polish, Vietnamese, German and Arabic Wikipedia. There also sources, that has relatively high position in all language versions: "variety.com" and deadline.com always in the top 20, "imdb.com" almost in all languages (except Japanese) in the top 20, 'who.int' in the top 100 of reliable sources in each considered languages.

6. Similarity of Models

According to the results presented in the previous section, each source can be placed on a different position in the ranking of the most reliable sources depending on the model. It is worthwhile to check how similar are the results obtained by different models. For this purpose we used Spearman's rank correlation to quantify, in a scale from −1 to 1 degree, which variables are associated. Initially we took only sources that appeared in the top 100 in at least one of the rankings of the most popular and reliable sources in multilingual Wikipedia in February 2020. Altogether, we obtained 180 sources and their positions in each of the rankings. Table 6 shows Spearman's correlation coefficients between these rankings.

Table 6. Spearman's correlation coefficients between rankings of the top 100 most popular and reliable sources in multilingual Wikipedia in February 2020 using different models.

Models	F	P	PR	PL	Pm	PmR	PmL	A	AR	AL
F	1.00	0.37	0.50	0.47	0.38	0.49	0.44	0.62	0.78	0.80
P	0.37	1.00	0.87	0.91	**0.99**	0.87	0.89	0.81	0.41	0.53
PR	0.50	0.87	1.00	**0.98**	0.87	1.00	0.94	0.82	0.61	0.68
PL	0.47	0.91	**0.98**	1.00	0.92	**0.97**	**0.97**	0.83	0.53	0.66
Pm	0.38	**0.99**	0.87	0.92	1.00	0.88	0.90	0.82	0.42	0.55
PmR	0.49	0.87	1.00	**0.97**	0.88	1.00	0.95	0.83	0.59	0.67
PmL	0.44	0.89	0.94	**0.97**	0.90	0.95	1.00	0.81	0.49	0.65
A	0.62	0.81	0.82	0.83	0.82	0.83	0.81	1.00	0.68	0.79
AR	0.78	0.41	0.61	0.53	0.42	0.59	0.49	0.68	1.00	0.92
AL	0.80	0.53	0.68	0.66	0.55	0.67	0.65	0.79	0.92	1.00

We can observe that the highest correlation is between rankings based on **P** and **Pm** model–0.99. This can be explained through similarities of the measures in models—the first is based on cumulative page views and the latter on median of daily page views in a given month.

Another pair of similar rankings is **PL** and **PR** models—0.98. Both measures use total page views data. In the first model value of this measure is divided by the number of references, in the second by article length. As we mentioned before in Sections 2 and 3, the number of references and article lengths are very important in quality assessment of the Wikipedia articles and are also correlated—we can expect that longer articles can have a bigger number of references.

In connection with previously described similarities between **P** and **Pm**, we can also explain similarity between models **PL** and **PmR** with 0.97 value of the Spearman's correlation coefficient.

The lowest similarity is between **F** and **P** model–0.37. It comes from different nature of these measures. In Wikipedia anyone can create and edit content. However, not every change in the Wikipedia articles can be checked by a specialist in the field, for example by checking reliability of the inserted sources in the references. Despite the fact that some sources are used frequently, there is a chance that they have not been verified yet and not replaced by more reliable sources. The next pair of rankings with the low correlation is **Pm** and **F** model. Such low correlation is obviously connected with similarity of the page view measures (**P** and **Pm**).

It is also important to note the low similarity between rankings based on **AR** and **P** models–0.41. Such differences can be connected with the measures that are used in these models. **AR** model uses the number of authors for whole edition history of article divided by the number of references whereas **P** uses page view data for selected month.

In the second iteration we extended the number of sources to top 10,000 in each ranking of the most popular and reliable sources in multilingual Wikipedia in February 2020. We obtained 19,029 sources. Table 7 shows Spearman's correlation coefficients between these extended rankings.

In case of extended rankings (top 10,000) there are no significant changes with regard to the the Spearman's correlation coefficient values compared to the top 100 model in Table 6. However, it should

be noted that the largest difference in values of coefficients appears between **PR** and **A** model–0.26 (0.82 in the top 100 and 0.56 in the top 10,000).

Table 7. Spearman's correlation coefficients between rankings of the top 10,000 most popular and reliable sources in multilingual Wikipedia in February 2020 using different models.

Models	F	P	PR	PL	Pm	PmR	PmL	A	AR	AL
F	1.00	0.38	0.51	0.50	0.38	0.49	0.46	0.68	0.80	0.84
P	0.38	1.00	0.67	0.79	0.99	0.67	0.78	0.72	0.37	0.47
PR	0.51	0.67	1.00	0.89	0.66	0.98	0.85	0.56	0.65	0.62
PL	0.50	0.79	0.89	1.00	0.78	0.88	0.96	0.59	0.51	0.62
Pm	0.38	0.99	0.66	0.78	1.00	0.69	0.79	0.73	0.37	0.48
PmR	0.49	0.67	0.98	0.88	0.69	1.00	0.88	0.57	0.64	0.62
PmL	0.46	0.78	0.85	0.96	0.79	0.88	1.00	0.59	0.49	0.62
A	0.68	0.72	0.56	0.59	0.73	0.57	0.59	1.00	0.72	0.81
AR	0.80	0.37	0.65	0.51	0.37	0.64	0.49	0.72	1.00	0.91
AL	0.84	0.47	0.62	0.62	0.48	0.62	0.62	0.81	0.91	1.00

The heatmap in Figure 5 shows Spearman's correlation coefficients between rankings of the top 100 most reliable sources in each language version of Wikipedia in February 2020 obtained by **F**-model in comparison with other models.

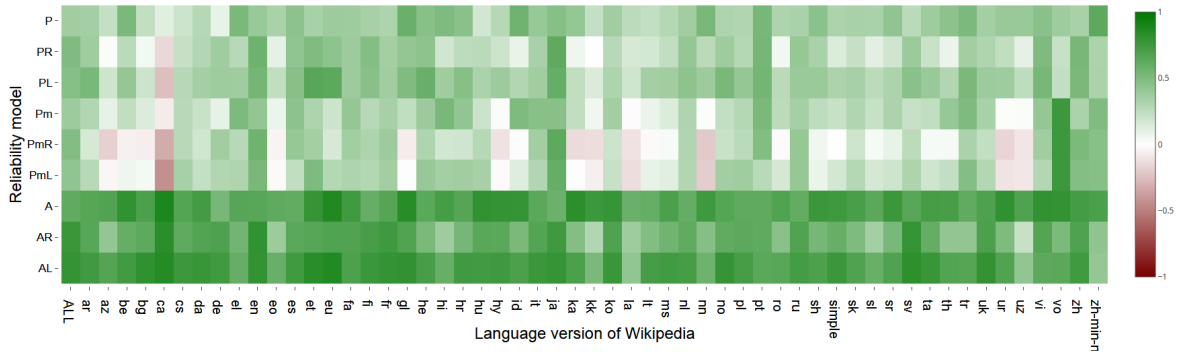

Figure 5. Spearman's correlation coefficients between rankings of the top 100 most reliable sources in each language version of Wikipedia in February 2020 obtained by **F**-model in comparison with other models. Interactive version of the heatmap is available on the web page: http://data.lewoniewski.info/sources/heatmap/.

Comparing the results of Spearman's correlation coefficients within each of considered language version of Wikipedia, we can find that the largest average correlation between **F**-model and other models is for Japanese (ja) and English (en) Wikipedia—0.61 and 0.59, respectively. The smallest average value of the correlation coefficients among languages have Catalan (ca) and Latin (la) Wikipedia—0.16 and 0.19, respectively. Considering coefficient values among all languages of each pair **F**-model and other model, the largest average value has **F**/**AL**-model pairs (0.71), the smallest—**F**/**PmR**-models (0.18).

7. Classification of Sources

7.1. Metadata from References

Based on citation templates in Wikipedia we are able to find more information about the source: authors, publication date, publisher and other. Using such metadata we decided to find which of the publishers and journals are most popular and reliable.

We first analyzed values of the publisher parameter in citations templates of the references of articles in English Wikipedia (as of March 2020). We found over 18 million references with citation

templates that have value in the publisher parameter. The Figure 6 shows the most commonly used publishers based on such analysis.

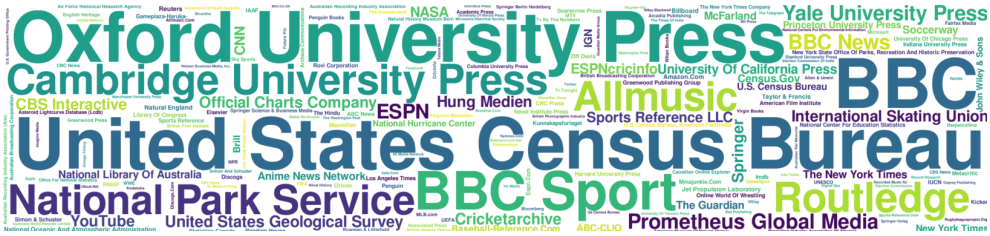

Figure 6. The most commonly used titles in publisher parameter of citations templates in the references of articles in English Wikipedia in March 2020. Source: own calculations based on Wikimedia dumps using complex extraction method.

Within the parameter publisher in references, the following names are most often found: United States Census Bureau, Oxford University Press, BBC, BBC Sport, Cambridge University Press, Routledge, National Park Service, AllMusic, Yale University Press, BBC News, Prometheus Global Media, United States Geological Survey, ESPN, CricketArchive, International Skating Union, Official Charts Company.

Using different popularity and reliability models we assessed all journals based on the related parameter in citation templates placed in references of English Wikipedia. Table 8 shows the most popular and reliable publishers with position in the ranking depending on the model.

Comparing the differences between ranking positions of the publishers using different models, we observed that some of the sources always have leading position: Oxford University Press (1st or 2nd place depending on model), BBC (2nd-5th place), Cambridge University Press (2nd-5th place), Routledge (3rd-6th place), BBC News (5th-10th place).

Some of the publisher has a high position in few models. For example, "United States Census Bureau" has the 1st place in **F** model (frequency) and **AR** model (authors per references count). At the same time in P (pageviews) model and **PL** model (pageviews per length of the text), this source took 27th and 11th places, respectively. Another one of the most frequent publisher in Wikipedia-'National Park Service' took 7th place. However it took only 94th and 58th places in **P** (pageviews) and **PmL** (pageviews median per length of the text) models, respectively. Publisher "Springer" took 5th place in **PmR** model (pageviews median per references count), but took only 19th place in F model (frequency). CNN took 2nd place in **P** (pageviews) and **Pm** (pageviews median) model, but at the same time took 22nd and 16th places in **F** (frequency) and **AR** (authors per references count) model, respectively. Wikimedia Foundation as a source in **P** (pageviews) model is in the top 10 sources, but at the same time is far from leading position in **F** (frequency) and **AR** (authors per length of the text) model—5541st and 3008th places, respectively.

It is important to note, that this ranking of publishers only take into account references with filled publisher parameter in citation templates in English Wikipedia, therefore it can not show complete information about leading sources in different languages (especially in those languages where citation templates are used rarely used).

Next we extracted values of journal parameter in citations templates of the references from articles in English Wikipedia. We found over 3 million references with citation templates that have value in the journal parameter. The Figure 7 shows the most commonly used journals based on such analysis. The most commonly used journals were: Nature, Astronomy and Astrophysics, Science, The Astrophysical Journal, Lloyd's List, PLOS ONE, Monthly Notices of The Royal Astronomical Society, The Astronomical Journal, Billboard.

Table 8. Position in rankings of publishers in English Wikipedia depending on popularity and reliability model in February 2020. Source: own calculation based on Wikimedia dumps using complex extraction and using only values from publisher parameter of citation templates in references. Extended version of the table is available on the web page: http://data.lewoniewski.info/sources/table8.

Source	Position in the Ranking Depending on Model									
	F	P	PR	PL	Pm	PmR	PmL	A	AR	AL
ABC News	71	18	54	39	20	57	43	34	83	61
ABC-CLIO	36	20	25	23	18	24	22	25	23	24
AllMusic	8	28	8	8	26	8	9	14	6	7
Anime News Network	27	34	12	15	32	14	15	40	20	30
BBC	3	4	3	3	5	4	3	2	3	3
BBC News	10	5	7	5	6	7	5	7	8	8
BBC Sport	4	11	15	12	16	17	13	5	7	5
Cambridge University Press	5	3	2	2	3	2	2	3	4	4
Canadian Online Explorer	119	85	199	127	84	198	124	13	66	29
CBS Interactive	20	9	10	7	9	10	8	12	15	10
CNN	22	2	9	6	2	9	6	6	16	12
CRC Press	53	63	20	27	58	20	25	63	40	45
Cricketarchive	14	823	460	445	980	736	655	210	74	79
ESPN	13	8	17	14	8	19	16	10	13	14
Harpercollins	121	17	52	36	14	48	32	28	65	55
Hung Medien	18	38	34	21	33	32	21	19	26	20
IGN	32	37	29	24	34	29	23	22	17	15
IMDB	65	55	16	17	65	16	19	89	34	47
International Skating Union	15	340	256	209	382	322	252	141	108	69
John Wiley & Sons	41	26	13	16	19	12	14	30	25	25
Macmillan	63	22	43	38	17	41	35	29	51	52
Metacritic	52	49	21	20	51	23	20	57	37	31
Microsoft	113	7	14	13	7	13	10	37	46	46
MTV	78	19	46	31	21	50	31	17	36	26
National Center For Education Statistics	66	2326	349	452	2431	396	495	226	11	17
National Park Service	7	94	38	48	89	47	58	60	12	11
Official Charts Company	16	30	24	18	31	26	18	18	21	18
Oxford University Press	2	1	1	1	1	1	1	1	2	1
Prometheus Global Media	11	36	27	22	37	30	27	20	18	16
Routledge	6	6	4	4	4	3	4	4	5	6
Simon & Schuster	75	14	26	25	11	27	26	23	45	41
Springer	19	12	6	9	10	5	7	16	10	13
The Hindu	72	183	18	26	163	15	24	185	56	65
United States Census Bureau	1	27	5	11	24	6	12	9	1	2
United States Geological Survey	12	170	78	95	159	80	98	74	14	21
University of California Press	24	16	19	19	12	18	17	15	19	22
Wikimedia Foundation	5541	10	145	92	47	379	251	1138	3008	2520
WWE	92	21	62	40	22	64	42	8	24	19
Yale University Press	9	13	28	28	13	28	29	21	33	28
YouTube	17	15	11	10	15	11	11	11	9	9

Figure 7. The most commonly used titles in the journal parameter of citations templates in the references of articles in English Wikipedia in March 2020. Source: own calculations based on Wikimedia dumps using complex extraction method.

Using different popularity and reliability models we assessed all journals based on the related parameter in citation templates placed in references of English Wikipedia. Table 9 shows the most popular and reliable journals with position in the ranking depending on the model. It is important to note that the same journal has two different names "Astronomy and Astrophysics" and "Astronomy & Astrophysics" because it was written in such ways in citation templates.

Table 9. Position in rankings of journals in English Wikipedia depending on popularity and reliability model in February 2020. Source: own calculation based on Wikimedia dumps using complex extraction and using only values from journal parameter in citation templates in references. Extended version of the table is available on the web page: http://data.lewoniewski.info/sources/table9.

Source	Position in Ranking Depending on Model Models									
	F	**P**	**PR**	**PL**	**Pm**	**PmR**	**PmL**	**A**	**AR**	**AL**
American Family Physician	84	36	42	20	25	38	17	20	47	19
Astronomy & Astrophysics	12	56	56	39	45	57	38	38	34	16
Astronomy and Astrophysics	2	31	25	11	22	25	12	12	7	4
Astronomy Letters	19	2085	1822	528	2311	2281	722	473	207	43
Billboard	9	16	8	9	12	7	8	6	5	7
BMJ	36	14	12	12	10	11	11	10	18	17
Cell	16	32	14	15	20	12	13	28	19	23
Communications of the ACM	188	29	3	4	38	17	36	119	54	99
Emory Law Journal	8049	11	114	77	37	480	302	2378	8573	6978
Icarus	14	21	38	27	16	36	25	11	20	14
JAMA	54	25	19	17	18	20	16	15	33	26
Journal of The American Chemical Society	30	79	21	29	52	18	27	61	28	38
Journal of Virology	120	33	18	24	24	19	23	233	203	199
Lancet	23	3	7	5	3	5	4	4	11	9
Lloyd's List	5	1278	5647	3196	2992	11528	8281	59	847	356
LPSN	17	4757	609	137	5978	1187	259	1820	94	36
Mammalian Species	56	77	67	42	58	66	39	31	36	20
MIT Technology Review	5565	5	57	41	19	209	132	1209	3900	3338
Molecular Phylogenetics and Evolution	34	101	41	48	94	43	46	47	17	21
Monthly Notices of The Royal Astronomical Society	7	30	26	19	21	26	21	18	13	8
Myconet	63	21506	640	1106	34191	2978	4134	3407	15	37
Nature	1	1	1	1	1	1	1	1	1	1
Nature News	885	20	110	85	48	228	200	406	410	522
New England Journal of Medicine	60	19	22	16	13	21	15	34	46	45
Pediatrics	62	38	43	35	28	40	32	16	39	28
Physical Review Letters	26	35	11	25	23	10	20	25	14	24
PLOS ONE	6	4	4	3	4	3	3	3	4	6
Proceedings of the National Academy of Sciences	18	15	9	10	11	8	9	9	9	15
Proceedings of the National Academy of Sciences of the United States of America	13	8	5	8	7	4	7	8	8	12
Rolling Stone	55	18	13	14	15	13	14	13	16	18
Science	3	2	2	2	2	2	2	2	2	2
The Astronomical Journal	8	42	58	33	35	63	34	24	21	11
The Astrophysical Journal	4	7	6	6	6	6	5	7	6	5
The Cochrane Database of Systematic Reviews	27	6	10	7	5	9	6	5	12	10
The Guardian	184	17	68	51	31	102	86	97	127	125
The IUCN Red List of Threatened Species	10	261	34	38	275	58	62	115	3	3
The Journal of American History	805	9	86	64	26	188	158	282	599	698
The Journal of Biological Chemistry	15	57	17	23	41	14	19	54	23	29
The Lancet	38	12	23	18	9	23	18	19	38	33
The New England Journal of Medicine	48	10	16	13	8	15	10	14	37	22
Time	64	13	20	26	14	24	26	17	25	27
Variety	86	34	15	22	27	16	24	37	26	35
Wired	141	22	30	28	17	30	28	26	52	51
Zookeys	20	649	193	172	734	295	289	362	42	42
Zootaxa	11	153	59	56	153	78	70	41	10	13

Comparing the differences between ranking positions of the journals using different models, we can also observe that some of the sources always have leading position: Nature (1st in all models), Science (2nd-3rd place depending on model), PLOS ONE (3rd-6th place), The Astrophysical Journal (4th-7th place).

Some of journals has a high position in few models. For example, "Lancet" journal took 3rd place in **P** (pageviews) and Pm (pageviews median) model, but is only on the 23rd place in **F** (frequency) model. Another example, "Proceedings of the National Academy of Sciences of the United States of America" has the 4th place in **PmR** model (pageviews median per references count) and at the same time 13th place in **F** (frequency) model. "Proceedings of The National Academy of Sciences" took 8th place in **PmR** model (pageviews per references count), but has 18th position in **F** model (frequency). There are journals that have signifficatly fidderent position depends on model. One of the good examples—"MIT Technology Review" that took 5th place in **P** model (pageviews), but only 5565th and 3900th places in **F** (frequency) and **AR** (authors count per references count) model, respectively.

Despite the fact that obtained results allow us to compare different meta data related to the source, we need to take into account significant limitation of this method-we can only assess the sources in references that used citation templates. Additionally, as we already discussed in Section 4.2, not always related parameters of the references are filled by Wikipedians. Therefore, we decided to take into account all references with URL address and conducted more complex analysis of the source types based on semantic databases.

7.2. Semantic Databases

Based on information about URL it is possible to identify title and other information related to the source. Using Wikidata [37,38] and DBpedia [39,40] we found over 900 thousand items (including such broadcasters, periodicals, web portals, publishers and other) which has aligned separate domain(s) or subdomain(s) as official site. Table 10 shows position in the global ranking of the most popular and reliable source with identified title based on found items in 55 considered language versions of Wikipedia in February 2020 using different models with identified title of the source

Leading positions in various models are occupied by following sources: Deadline Hollywood, TV by the Numbers, Variety, Internet Movie Database. "Forbes", "The Washington Post", "CNN", "Entertainment Weekly", "Oricon" are in the top 20 of all rankings in Table 10. We can also observe sources with relative big differences in rankings between the models. For example, "Newspapers" (historic newspaper archive) in on the 5th place of the most frequent used sources in Wikipedia, at the same time is on 33rd and 23rd place in **Pm** (pageviews median) and **PmL** (pageviews median per length of the text) models respectively. Another example, "Soccerway" is on the 7th place in the ranking of the most commonly used sources (based on **F** model), but is on 116th and 100th places in **P** and **Pm** models, respectively. Despite the fact that "American Museum of Natural History" is on top 20 the most commonly used sources in Wikipedia (based on F model), it is excluded from top 5000 in **P** (pageviews), **Pm** (pageviews median), PmR (pageviews median per reference count) and PmL ((pageviews) median per length of text) models.

Table 11 shows the most popular and reliable types of the sources in selected language versions of Wikipedia in February 2020 based on **PR** model. In almost all language versions websites are the most reliable sources. Magazines and business related source are top 10 of the most reliable types of sources in all languages. Film databases are one of the most reliable sources in Arabic, French, Italian, Polish and Portuguese Wikipedia. In other languages such sources are placed above 19th place. Arabic, English, French, Italian and Chinese Wikipedia preferred newspapers as a reliable source more than in other languages that placed such sources lower in the ranking (but above the 14th place). News agencies are more reliable for Persian Wikipedia comparing with other languages. Government agencies as a source has much more reliability in Persian and Swedish Wikipedia than in other languages. Holding companies provides more reliable information for Japanese and Chinese languages. In Dutch and Polish Wikipedia archive websites has relatively higher position in the reliability ranking. Periodical sources are more reliable German, Spanish and Polish Wikipedia. Review aggregators are more reliable in Arabic and Polish Wikipedia comparing other considered languages. Television networks in on 7th place in German Wikipedia and on 14th place in Portuguese Wikipedia, while other languages have such sources even on lower then 20th place (even 125th place).

Social networking services are placed in top 20 of the most reliable types of sources in Japanese, Polish and Chinese Wikipedia. Weekly magazines are in the top 10 of English, Italian, Portuguese and Russian Wikipedia.

Table 10. Position in the global ranking of the most popular and reliable sources with identified title in 55 considered language versions of Wikipedia depending on the model in February 2020. Source: own calculations based on Wikimedia dumps using complex extraction of references. Extended version of the table is available on the web page: http://data.lewoniewski.info/sources/table10.

Source	Model									
	F	**P**	**PR**	**PL**	**Pm**	**PmR**	**PmL**	**A**	**AR**	**AL.**
American Museum of Natural History	19	6048	685	946	6941	5880	6984	459	7	9
CBS News	42	13	33	33	13	36	35	23	49	44
CNN	14	7	17	15	7	16	17	4	14	15
Collider	55	16	27	25	15	27	22	39	65	57
Deadline Hollywood	1	1	1	1	1	1	1	1	3	3
Entertainment Weekly	12	5	5	6	5	5	6	5	13	14
Forbes	20	8	10	8	8	9	7	15	20	17
GameSpot	11	19	24	14	14	24	14	11	15	12
IndieWire	81	15	16	17	19	19	21	82	109	102
Internet Movie Database	4	21	3	5	21	3	4	6	1	1
MTV	21	18	29	29	17	29	29	7	16	18
Newspapers.com	5	30	15	20	33	20	23	17	11	7
Official Charts	10	31	20	22	28	17	18	13	12	10
Oricon	9	11	7	4	11	7	5	12	8	5
People	53	17	12	11	20	11	12	22	23	23
Pitchfork	15	29	23	18	27	21	16	21	18	16
Rotten Tomatoes	17	10	6	7	10	6	8	18	9	13
Soccerway	7	100	40	52	116	50	60	32	10	11
TV by the Numbers	2	3	4	3	3	4	3	3	6	8
TVLine	43	26	18	27	24	15	26	45	47	52
TechCrunch	34	20	26	13	16	22	11	38	52	32
The Atlantic	48	12	35	34	18	37	37	33	53	46
The Daily Telegraph	28	14	21	21	12	25	20	20	31	25
The Futon Critic	18	36	19	30	35	18	30	27	27	36
The Indian Express	31	37	14	16	36	13	15	25	26	22
The Washington Post	13	4	9	9	4	10	9	8	17	19
Time	29	9	22	19	9	23	19	19	33	27
USA Today	16	6	11	10	6	12	10	10	19	20
Variety	3	2	2	2	2	2	2	2	2	2
Wayback Machine	8	38	13	24	37	14	25	16	5	6
WordPress.com	6	33	8	12	31	8	13	9	4	4

Based on the knowledge about type of each source we decided to limit the ranking to specific area. We chosen only periodical sources which aligned to one of the following types: online newspaper (Q1153191), magazine (Q41298), daily newspaper (Q1110794), newspaper (Q11032), periodical (Q1002697), weekly magazine (Q12340140). The top of the most reliable periodical sources in all considered language versions in Wikipedia in February 2020 occupies: Variety, Entertainment Weekly, The Washington Post, USA Today, People, The Indian Express, The Daily Telegraph, Time, Pitchfork, Rolling Stone.

The most popular periodical sources in Wikipedia articles from 55 language versions using different popularity and reliability models in February 2020 showed in Table 12. There are sources that have stable reliability in all models–"Variety" has always 1st place, "Entertainment Weekly" 2nd-3nd place, "The Washington Post" occupies 2nd-4th place, "USA Today" took 4th-5th place depending on the model. Despite the fact that "Lenta.ru" is the 6th most commonly used periodical source in different languages of Wikipedia (using **F** model), it is placed in 21st and 19th places using **P** and **Pm** models, respectively. "The Daily Telegraph" is in the top 10 most reliable periodical sources in all

models. "People" is in 18th place in frequency ranking, but at the same time took 4th place in the **PmR** model.

Table 11. The most popular and reliable types of the sources in selected language versions of Wikipedia in February 2020 based on **PR** model. Source: own calculations based on Wikimedia dumps using complex extraction of references with semantic databases (Wikidata, DBpedia) to identify type of the source. Extended version of the table is available on the web page: http://data.lewoniewski.info/sources/table11.

Source type	Language Version of Wikipedia														
	ar	**de**	**en**	**es**	**fa**	**fr**	**it**	**ja**	**nl**	**pl**	**pt**	**ru**	**sv**	**vi**	**zh**
archive	12	56	39	12	27	30	24	21	3	6	31	38	36	21	58
business	7	3	5	5	2	6	9	3	7	3	2	5	5	3	3
daily newspaper	9	4	4	6	10	8	2	4	1	16	5	4	6	9	9
enterprise	14	6	7	8	6	10	8	6	9	7	8	6	7	5	4
film database	2	10	10	9	7	3	5	5	13	2	4	8	18	17	10
government agency	25	75	51	60	4	52	59	45	71	24	62	60	4	62	56
holding company	135	252	62	133	194	115	152	2	471	99	98	141	391	35	7
magazine	8	2	2	2	5	7	4	7	4	5	3	2	8	4	6
morning paper	164	245	221	544	445	387	501	644	417	505	482	540	2	381	504
natural history museum	561	583	391	579	800	405	442	792	19	478	414	510	556	10	523
news agency	40	113	49	65	3	61	56	72	114	104	99	66	124	54	53
news website	21	12	6	4	13	9	7	15	17	20	6	7	42	15	5
newspaper	3	8	3	7	9	2	3	9	5	13	7	9	9	7	2
online database	4	13	12	14	11	5	13	41	12	10	15	12	17	16	26
online newspaper	18	26	13	10	24	20	23	23	23	33	25	3	37	2	12
open-access publisher	17	18	26	20	18	19	22	30	26	25	19	32	26	8	17
organization	11	9	9	11	8	4	11	10	10	9	9	11	10	6	13
periodical	37	5	15	3	22	11	12	36	6	4	22	13	12	12	34
public broadcasting	66	80	36	77	52	78	82	18	8	77	87	93	112	45	80
review aggregator	6	15	16	17	15	14	16	25	15	8	21	19	20	24	27
social cataloging application	5	14	14	15	14	13	14	48	14	12	20	18	19	23	30
social networking service	33	30	22	29	26	27	29	16	28	14	35	31	59	30	8
specialty channel	10	23	11	22	17	18	18	34	21	15	17	17	28	18	18
television network	53	7	35	33	45	38	77	125	82	87	14	84	21	37	79
television station	20	16	17	27	20	22	19	8	16	21	29	14	54	19	20
website	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1
weekly magazine	26	11	8	16	16	12	10	24	18	23	10	10	41	13	11
written work	123	256	167	104	430	64	6	529	141	155	78	164	418	263	519

Given local rankings of periodical we can consider the difference of reliability and popularity between different language versions. Table A2 shows the position in local rankings of periodical sources in different language versions of Wikipedia in February 2020 using **PR** model. Almost in all considered languages (except Dutch) "Variety" took 1st-4th places in local rankings of the most reliable periodical sources. Some sources that are in leading positions in local rankings are not presentet at all as a sources in some languages. For example. "Aliqtisadi" (Arabic news magazine) is in the 2nd place in Arabic Wikipedia, but in English, Persian, Italian, Japanese, Russian Wikipedia position this source is lower then 600th place and not presented in other language as a source. Similar tendencies is to "Ennahar newspaper", which has 5th place in Arabic Wikipedia. For the German Wikipedia 2nd, 3rd and 4th place belongs to "Die Tageszeitung", "DWDL.de", "Auto, Motor und Sport". For Spanish Wikipedia leading local periodical sources are: "20 minutos", "El Confidencial", "Entertainment Weekly", "¡Hola!". In Persian Wikipedia one of the most reliable periodical source "Donya-e-Eqtesad", that is not presented at all in most of the considered languages. The most reliable sources in French Wikipedia include: "Le Monde", "Jeune Afrique", "Le Figaro", "Huffington Post France". Italian version of Wikipedia contains such the most reliable local sources as: "la Repubblica", "Il Post", "Il Fatto Quotidiano". In Japan Wikipedia leading reliable sources includes "Nihon Keizai Shimbun", "Tokyo Sports", "Yomiuri Shimbun". Dutch Wikipedia contains "De Volkskrant", "Algemeen Dagblad", "Het Laatste Nieuws", "Trouw", "NRC Next" as one of the most reliable periodical sources. Polish Wikipedia has "Wprost" and "TV Guide" in top 3 periodical sources. In Portuguese one of the most reliable periodical sources are "Veja" and "Exame". "Lenta.ru"

and "Komsomolskaya Pravda" are leading periodical sources in Russian Wikipedia. Swedish language version has "Sydsvenskan", "Dagens Industri" and "Helsingborgs Dagblad" as leading reliable sources. "VnExpress" took 1st place in the most reliable periodical sources of Vietnamese Wikipedia. "Apple Daily" is the most reliable periodical source in Chinese language version.

Table 12. The most popular periodical sources in Wikipedia articles from 55 language versions using different popularity and reliability models in February 2020. Source: own calculations based on Wikimedia dumps using complex extraction of references with semantic databases (Wikidata, DBpedia) to identify type of the source. Extended version of the table is available on the web page: http://data.lewoniewski.info/sources/table12.

Source	Models									
	F	**P**	**PR**	**PL**	**Pm**	**PmR**	**PmL**	**A**	**AR**	**AL**
Entertainment Weekly	2	3	2	2	3	2	2	2	2	2
Flight International	20	19	25	22	17	22	20	17	26	20
Fortune	36	15	17	17	15	17	16	25	36	28
Komsomolskaya Pravda	21	36	24	28	37	23	29	31	20	26
Lenta.ru	6	21	13	16	19	13	17	14	9	11
New York Post	27	18	21	18	16	20	21	19	24	21
Nihon Keizai Shimbun	14	27	16	13	26	16	13	24	16	15
PC Gamer	28	25	22	20	24	21	18	26	35	30
People	18	8	5	5	9	4	6	8	6	7
Pitchfork	4	13	9	8	12	7	8	7	4	3
Rolling Stone	16	11	10	11	10	11	11	10	15	16
Spin	26	29	30	30	30	30	31	20	29	22
TV Guide	33	28	18	21	27	19	22	29	19	23
TechCrunch	11	9	11	6	7	8	5	16	17	13
Technology Review	107	16	48	41	28	61	52	95	118	116
The Atlantic	17	6	14	14	8	15	15	13	18	18
The Daily Telegraph	7	7	7	10	6	10	10	6	10	8
The Express Tribune	24	42	28	26	40	26	25	27	23	19
The Globe and Mail	10	22	19	19	20	18	19	15	12	14
The Indian Express	9	14	6	7	14	6	7	9	7	6
The Japan Times	42	23	32	32	18	32	35	45	43	43
The New York Times	12	12	15	15	13	14	14	12	13	12
The Wall Street Journal	29	20	27	25	22	28	27	23	31	27
The Washington Post	3	2	3	3	2	3	3	3	3	4
Time	8	5	8	9	5	9	9	5	11	9
USA Today	5	4	4	4	4	5	4	4	5	5
Ukrayinska Pravda	19	61	76	68	61	76	72	35	49	42
Variety	1	1	1	1	1	1	1	1	1	1
Wired	13	10	12	12	11	12	12	11	14	10
la Repubblica	15	17	20	24	21	24	30	18	8	17

8. Temporal Analysis

Using complex extraction of the references apart from data from February 2020, we also used dumps from November 2019, December 2019 and January 2020. Based on those data we measure popularity and reliability of the sources in different months.

Table 13 shows position in rankings of popular and reliability sources with identified title depending on period in all considered languages versions of Wikipedia using **PR** model. Results showed that some of the sources didn't changes their position in the ranking based on **PR** model. This is especially applicable to sources with leading position. For example "Deadline Hollywood", "Variety", "Entertainment Weekly", "Rotten Tomatoes", "Oricon" in each of the studied month he occupied the same place in top 10. "Internet Movie Database" and "TV by the Numbers" exchanged 3rd and 4th places. This is due to the fact that in absolute values of popularity and reliability measurement obtained using **PR** model, most of these sources have significant breaks from the closest competitors.

Table 13. Position in rankings of popular and reliable sources depending on period in all considered language versions of Wikipedia using **PR** model. Source: own work based on Wikimedia dumps using complex extraction of references with semantic databases (Wikidata, DBpedia) to identify title of the sources. Extended version of the table is available on the web page: http://data.lewoniewski.info/sources/table13.

Sources	Months			
	December 2019	January 2020	February 2020	March 2020
CNN	18	20	16	17
Deadline Hollywood	1	1	1	1
Entertainment Weekly	5	5	5	5
Forbes	9	9	9	10
GameSpot	17	16	22	24
IndieWire	24	17	20	16
Internet Movie Database	4	3	4	3
Newspapers.com	19	18	18	15
Official Charts	15	19	21	20
Oricon	7	7	7	7
People	12	10	11	12
Rotten Tomatoes	6	6	6	6
TV by the Numbers	3	4	3	4
TVLine	14	15	14	18
The Daily Telegraph	20	21	17	21
The Futon Critic	21	23	19	19
The Indian Express	16	12	15	14
The Washington Post	11	14	12	9
USA Today	13	11	10	11
Variety	2	2	2	2
Wayback Machine	10	13	13	13
WordPress.com	8	8	8	8

Next we decided to limit the list of the sources to periodical ones (as it was done in Section 7.2). Table 14 shows position in rankings of popular and reliable sources depending on period in all considered languages versions using **PR** model. Similarly to the previous table, we can observe not significant changes in position for the leading sources. In four considered months the top 10 most reliable periodical sources always included: "Variety", "Entertainment Weekly", "The Washington Post", "People", "USA Today", "The Indian Express", "The Daily Telegraph" "Pitchfork", "Time".

Results showed, that in the case of periodical sources we have less "stability" of the position in the ranking between different months comparing to the general ranking. For reasons already explained, the 2 top sources (Variety and Entertainment Weekly) did not change their positions. Additionally we can distinguish The Daily Telegraph with stable 7th place during whole considered period of time. Nevertheless in top 10 the most popular and reliable periodical sources of Wikipedia we can observe minor changes in positions. This applies in particular to People, Pitchfork, The Washington Post, USA Today, The Indian Express, Time. those sources grew or fell by 1-2 positions in the top 10 ranking during the November 2019-February 2020.

As it was mentioned before, minor changes in the ranking of sources during the considered period are mainly due to a large margin in absolute values of popularity and reliability measurement. This applies in particular to leading sources. However, what if there are relatively new sources that have significant prerequisites to be leaders or even outsiders in nearest future. The next section will describe the method and results of measuring.

Table 14. Position in rankings of popular and reliable sources depending on period in all considered language versions using **PR** model. Source: own work based on Wikimedia dumps using complex extraction of references with semantic databases (Wikidata, DBpedia) to identify type of the source. Extended version of the table is available on the web page: http://data.lewoniewski.info/sources/table14.

Sources	Months			
	December 2019	January 2020	February 2020	March 2020
Apple Daily	29	31	30	35
Empire	32	29	33	33
Entertainment Weekly	2	2	2	2
Flight International	23	24	20	25
Fortune	17	19	19	17
GamesMaster	28	28	29	29
Komsomolskaya Pravda	21	22	23	24
la Repubblica	25	25	25	20
Lenta.ru	11	12	12	13
Metro	24	23	26	23
New York Post	20	21	21	21
Nihon Keizai Shimbun	15	15	16	16
PC Gamer	22	20	24	22
People	4	3	4	5
Pitchfork	8	9	9	9
Radio Times	26	26	22	26
Rolling Stone	12	11	10	10
Spin	30	32	32	30
TV Guide	19	17	18	18
TechCrunch	9	10	11	11
The Atlantic	16	16	15	14
The Daily Telegraph	7	7	7	7
The Express Tribune	37	30	27	28
The Globe and Mail	18	18	17	19
The Indian Express	6	5	6	6
The New York Times	14	14	14	15
The Wall Street Journal	27	27	28	27
The Washington Post	3	6	5	3
Time	10	8	8	8
USA Today	5	4	3	4
Variety	1	1	1	1
Wired	13	13	13	12

9. Growth Leaders

The Wikipedia articles may have a long edition history. Information and sources in such articles can be changed many times. Moreover, criteria for reliability assessment of the sources can be changed over time in each language version of Wikipedia. Based on the assessment of the popularity and reliability of each source in Wikipedia in certain period of time (month) we can compare the differences between the values of the measurement. This can help to find out how popularity and reliability were changed (increase or decrease) in a particular month. For example, a certain Internet resource has only recently appeared and people have actively begun to use it as a source of information in Wikipedia articles. Another example: a well known and often used website in Wikipedia references dramatically lost confidence (reputation) as a reliable source, and editors actively start to replace this source with another or place additional reference next to existing ones. First place in such ranking means, that for the selected source we observed the largest growth of the popularity and readability score comparing previous month.

Table 15 shows which of the periodical sources had the largest growth of reliability in selected languages and period of times based on **F** model. For this table we have chosen only sources which was placed at least in top 5 in the growth leaders ranking of the one of the languages and selected month. Results shows that there is no stable growth leaders for the sources when we comparing different periods of time.

F model showed how many references in Wikipedia articles contain specific sources. Therefore, we can analyze which of the sources was more often added in references in Wikipedia articles in the considered month. For example in December 2019 "Die Tageszeitung" and "Handelsblatt" were leading growing sources in German Wikipedia, "Jeune Afrique" and "Les Inrockuptibles" were leading growing sources in French Wikipedia, "Komsomolskaya Pravda" and "Lenta.ru" were leading growing sources in Russian Wikipedia. In next month (January 2020) "Süddeutsche Zeitung" and "Die Tageszeitung" were leading growing sources in German Wikipedia, "Variety" and "La Montagne" were leading growing sources in French Wikipedia, "Variety" and "Komsomolskaya Pravda" were leading growing sources in Russian Wikipedia. In the last considered month (February 2020) "Die Tageszeitung" and "Variety" were leading growing sources in German Wikipedia, "Jeune Afrique" and "La Montagne" were leading growing sources in French Wikipedia, "Sport Express" and "Variety" were leading growing sources in Russian Wikipedia.

Table 15. Position of the periodical sources in growth ranking in selected language versions of Wikipedia and period of time using **F** model. Source: own work based on Wikimedia dumps using complex extraction of references with semantic databases (Wikidata, DBpedia). Extended version of the table is available on the web page: http://data.lewoniewski.info/sources/table15.

Source	German Wikipedia (de)			French Wikipedia			Russian Wikipedia		
	December 2019	January 2020	February 2020	December 2019	January 2020	February 2020	December 2019	January 2020	February 2020
Auto, Motor und Sport	14	18	4	2326	2341	2373	1007	1033	82
Daily Herald	505	3103	5	623	673	691	659	686	698
Die Tageszeitung	1	2	1	108	97	67	110	2715	185
El Observador	363	3	3280	836	882	901	583	621	625
Entertainment Weekly	10	49	34	10	39	11	17	3	11
GamesMaster	76	86	66	101	110	5	10	8	22
Handelsblatt	2	13	3269	1743	1764	1799	2517	2535	2571
Jeune Afrique	59	270	40	1	3	1	163	202	124
Jüdische Allgemeine	4	20	3	372	1120	1145	998	1024	1051
Komsomolskaya Pravda	106	339	2612	125	135	140	1	2	3
La Montagne	1919	749	1289	4	2	2	-	-	-
Lenta.ru	317	73	3159	252	177	78	2	5	5
Les Inrockuptibles	183	153	2619	2	5	3	124	79	398
Metal.de	27	5	3278	164	254	480	327	396	127
News.de	35	4	3279	1406	1433	165	938	964	989
Objectif Gard	1292	1503	1025	83	4	13	-	-	-
Pitchfork	42	42	50	13	38	21	5	6	4
Sport Express	179	187	2946	44	94	79	7	4	1
Süddeutsche Zeitung	3076	1	3281	285	573	588	383	445	422
TVyNovelas	2765	2806	2341	886	374	912	4	399	369
The Washington Post	13	29	6	5	11	9	14	16	13
Time	5	30	35	20	28	23	15	20	21
Variety	3	6	2	3	1	4	3	1	2

Table 16 shows which of the sources had the largest growth of reliability in different languages and period of times based on **PR** model. For this table we also have chosen only sources which was placed at least in top 5 in the growth leaders ranking of the one of the languages and selected month. Results showed also that there is no stable growth leaders for the sources when we comparing different period of time.

PR model showed how many references in Wikipedia articles contains specific sources with taking into account popularity of the articles. Results showed that in December 2019 Variety and Deutsche Jagd-Zeitung were leading growing reliable sources in German Wikipedia, Variety and Entertainment Weekly were leading growing reliable sources in French Wikipedia, "Lenta.ru" and Entertainment Weekly were leading growing sources in Russian Wikipedia. In next month (January 2020) "Die Tageszeitung" and "DWDL.de" were leading growing sources in German Wikipedia, "Les Inrockuptibles" and "Le Monde" were leading growing sources in French Wikipedia, Variety and "Lenta.ru" were leading growing sources in Russian Wikipedia. In the last considered month (February 2020) "la Repubblica" and "Algemeen Dagblad" were leading growing sources in German Wikipedia, "Atlanta" (magazine) and "Le Figaro étudiant" were leading growing sources in French Wikipedia, New York Post and "Novosti Kosmonavtiki" were leading growing sources in Russian Wikipedia.

Table 16. Position of the sources in growth ranking in selected language versions of Wikipedia and period of time using **PR** model. Source: own work based on Wikimedia dumps using complex extraction of references with semantic databases (Wikidata, DBpedia). Extended version of the table is available on the web page: http://data.lewoniewski.info/sources/table16.

Source	German Wikipedia (de)			French Wikipedia			Russian Wikipedia		
	2019-12	2020-01	2020-02	2019-12	2020-01	2020-02	2019-12	2020-01	2020-02
Algemeen Dagblad	53	2941	2	3039	3000	162	2236	2518	157
Atlanta (magazine)	3020	3090	3214	19	3152	1	244	2657	2274
Auto, Motor und Sport	3075	5	3	2688	446	2553	2330	350	150
Deutsche Jagd-Zeitung	2	3103	3130	-	-	-	-	-	-
Die Tageszeitung	3076	1	3276	3044	102	3011	2509	116	2694
DWDL.de	3073	2	3280	2200	279	12	2128	2317	1482
Entertainment Weekly	3	3102	3275	2	3151	3180	2	13	2777
Izvestia	195	2894	2630	953	631	2306	19	21	5
Jeune Afrique	2741	216	3024	3	9	3177	166	670	2161
Komsomolskaya Pravda	168	2618	3088	3016	3021	14	7	3	2779
la Repubblica	99	56	1	3113	12	3162	2615	9	2767
Le Figaro étudiant	778	1889	1516	3080	26	2	691	2311	1861
Le Monde	171	433	3067	3121	2	3178	2564	449	2473
Lenta.ru	19	3071	3224	397	3016	16	1	2	2780
Les Inrockuptibles	2739	264	2918	3122	1	3172	390	2494	2316
New York Post	2984	81	3221	42	22	3158	103	25	1
Novosti Kosmonavtiki	563	1624	457	2276	807	2468	2657	2689	2
PC Gamer	3043	173	3213	3053	164	3145	5	10	2774
People	3051	8	3266	3101	5	3171	2733	4	2772
Politico	100	2579	4	715	1267	3	216	239	231
Polka Magazine	-	-	-	1255	876	5	-	-	-
Radio Times	60	20	3259	4	62	3161	2723	8	2766
Russkij medicinskij zhurnal	1773	2958	2709	-	-	-	143	5	2768
Sankt-Peterburgskie Vedomosti	698	1838	914	899	2315	1255	2728	40	3
Sport Express	2889	850	2647	2978	3018	2345	3	2748	2776
Süddeutsche Zeitung	3064	4	3281	3012	283	2967	2590	125	2673
The Daily Gazette	1351	698	2629	54	3117	4	2132	680	1986
The Daily Telegraph	4	3076	3267	13	25	3163	352	2474	1365
The Tennessean	2734	164	5	153	97	31	2474	223	2614
Time	5	3099	3273	5	19	3176	4	94	2773
USA Today	18	10	3268	16	3	3166	16	31	2762
Variety	1	3	3269	1	4	3181	9	1	2778
Vedomosti	341	633	2764	2396	1286	1897	24	2735	4

10. Discussion of the Results

This study describes different models for popularity and reliability assessment of the sources in different language version of Wikipedia. In order to use these models it is necessary to extract information about the sources from references and also measures related to quality and popularity of the Wikipedia articles. We observed that depending on the model positions of the websites in the rankings of the most reliable sources can be different. In language versions that are mostly used on the territory of one country (for example Polish, Ukrainian, Belarusian), the highest positions in such rankings are often occupied by local (national) sources. Therefore, community of editors in each language version of Wikipedia can have own preferences when a decision is made to enable (or disable) the source in references as a confirmation of the certain fact. So, the same source can be reliable in one language version of Wikipedia, while the community of editors of another language may not accept it in the references and remove or replace this source in an article.

The simplest of the proposed models in this study was based on frequency of occurrences, which is commonly used in related studies. Other 9 novel models used various combinations of measures related to quality and popularity of Wikipedia articles. We provided analysis on how the results differ depending on the model. For example, if we compare frequency-based (F) rankings with other (novel) in each language version of Wikipedia, then the highest average similarity will have **AL**-model (0.71 of rank correlation coefficient), the least – **PmR**-model (0.18 of rank correlation coefficient).

The analysis of sources was conducted in various ways. One of the approaches was to extract information from citation templates. Based on the related parameter in references of English Wikipedia we found the most popular publishers (such as United States Census Bureau, Oxford University

Press, BBC, Cambridge University Press). The most commonly used journals in citation templates were: Nature, Astronomy and Astrophysics, Science, The Astrophysical Journal, Lloyd's List, PLOS ONE, Monthly Notices of The Royal Astronomical Society, The Astronomical Journal, Billboard. However, such approach was limited and did not include references without citation templates. Therefore, we decided to use semantic databases to identify the sources and their types.

After obtaining data about types of the sources we found that magazines and business-related sources are in the top 10 of the most reliable types of sources in all considered languages. However, the preferred type of source in references depends on language version of Wikipedia. For example, film databases are one of the most reliable sources in Arabic, French, Italian, Polish and Portuguese Wikipedia. In other languages such sources are placed below 19th place.

Including data from Wikidata and DBpedia allowed us to find the best sources in specific area. Using information about the source types and after choosing only periodical ones, we found that there are sources that have stable reliability in all models - "Variety" has always 1st place, "Entertainment Weekly" 2nd-3nd place, "The Washington Post" occupies 2nd-4th place, "USA Today" took 4th-5th place depending on the model. Despite the fact that "Lenta.ru" is the 6th most commonly used periodical source in different languages of Wikipedia (using **F** model), it is placed on 21st and 19th place using **P** and **Pm** models respectively. "The Daily Telegraph" is in the top 10 most reliable periodical sources in all models. "People" is on 18th place in the frequency ranking but at the same time took 4th place in **PmR** model.

Using complex extraction of the references in addition to data from February 2020 we also used dumps from November 2019, December 2019, and January 2020. Based on those data we measured popularity and reliability of the sources in different months. After limiting the sources to periodicals we found that in four considered months the top 10 most reliable periodical sources in multilingual Wikipedia always included: "Variety", "Entertainment Weekly", "The Washington Post", "People", "USA Today", "The Indian Express", "The Daily Telegraph", "Pitchfork", and "Time". Minor changes in the ranking of sources appearing during the considered period are mainly due to a large margin in absolute values of popularity and reliability measurement.

Different approaches assessing reliability of the sources presented in this research contribute to a better understanding which references are more suitable for specific statements that describe subjects in a given language. Unified assessment of the sources can help in finding data of the best quality for cross-language data fusion. Such tools as DBpedia FlexiFusion or GlobalFactSync Data Browser [41,42] collect information from Wikipedia articles in different languages and present statements in a unified form. However, due to independence of edition process in each language version, the same subjects can have similar statements with various values. For example, population of the city in one language can be several years old, while other language version of the article about the same city can update this value several times a year on a regular basis along with information about the source. Therefore, we plan to create methods for assessing sources of such conflict statements in Wikipedia, Wikidata and DBpedia to choose the best one. This can help to improve quality in cross-language data fusion approaches.

Proposed models can also help to assess the reliability of sources in Wikipedia on a regular basis. It can support understanding preferences of the editors and readers of Wikipedia in particular month. Additionally, it can be helpful to automatically detect sources with low reliability before user will insert it in the Wikipedia article. Moreover, results obtained using the proposed models may be used to suggest Wikipedians sources with higher reliability scores in selected language version or selected topic.

10.1. Effectiveness of Models

In this section we present the assessment of the models' effectiveness. Python algorithms prepared for purposes of this study were tested on desktop computer with Intel Core i7-5820K CPU and SSD

hard drive. Algorithms used only one thread of the processor. Due to the fact that each model used own set of measures, we divided assessment into several stages, including extracting of:

- External links using basic extraction method on compressed gzip dumps with total volume 12 GB-0.28 milliseconds per article on average.
- Sources from references using complex extraction method on bzip2 dumps with total volume 64 GB-2 milliseconds per article on average.
- Text length of articles (as a number of characters) using compressed bzip2 dumps with total volume 64 GB-0.68 milliseconds per article on average.
- Total page views for considered month using compressed bzip2 dumps with total volume 12 GB-0.25 milliseconds per article on average.
- Median of daily page views for considered month using compressed bzip2 dumps with total volume 12 GB-0.26 milliseconds per article on average.
- Number of authors of articles using compressed bzip2 dumps with total volume 170 GB-1.12 milliseconds per article on average.

Given the above and the fact we can calculate the effectiveness for each model during conversion, time the algorithm needs to calculate the popularity and reliability of the source is as follows:

- **F** model: 2 milliseconds per article.
- **P**, **PR** model: 2.25 milliseconds per article.
- **Pm**, **PmR** model: 2.28 milliseconds per article.
- **PL** model: 2.93 milliseconds per article.
- **PmL** model: 2.94 milliseconds per article.
- **A**, **AR** model: 3.12 milliseconds per article.
- **AL** model: 3.8 milliseconds per article.

10.2. Limitations

Reliability as one of the quality dimensions is a subjective concept. Each person can have their own criteria to asses reliability of the gives sources. Therefore each Wikipedia language community can have its own definition of reliable source. Only English Wikipedia, as the most developed edition of this free encyclopedia, provided an extended list of reliable/unreliable sources [43]. However it not always been used-for example despite the fact that IMDb (Internet Movie Database) is market as 'Generally unreliable' it is used very often (see Figure 4 or Table A1). As we observed, in some cases such sources can be used in references with some limitations—it can describe some specific statements (but not all). Therefore additional analysis of the placement of such sources in the articles can help to find such limited areas, where some sources can be used.

In the study we proposed and used 10 models to assess the popularity and reliability of the sources in Wikipedia. Each of the model use some of the important measures related to content popularity and quality. However, there are other measures that have potential to improve presented approach. Therefore we plan to extend the number of such measures in model. We plan to analyze possibility of comparing the results with other approaches or lists of the sources. For example it can be the most popular websites based on special tools, or reliable sources according to selected standards in some countries.

Each of the model can have own weak and strong sides. For example, during the experiments we observed, that some of articles has overstated values of the page views in some languages in selected months. This can be deduced from other related measures of the article. Sources in such articles could get extra points. However, these were individual cases that did not significantly affect the results of the work. In future work we plan to provide additional algorithms to automatically find and reduce such cases.

To extract the sources from references, which usually are published as of the first day of each month. We have information only for specified timestamp of the articles and we do not analyze in what day the source was inserted (or deleted) in the Wikipedia article. If the source was inserted few minutes (seconds) before the process of creating dumps files was started, we will count it as it was presented during the last considered month. Moreover, it can be more negatively involve on the model if such source was deleted few minutes (seconds) after the dump creating was begun. In other words, if the reference with the specified source was inserted and deleted around the timestamp of dump files creation, it can slightly or strongly (depend on values of article measures) falsify the results of some of the models. Therefore, more detailed analysis of each edition of the article can help to find how long particular reference was presented in article.

11. Conclusions and Future Work

In this paper we used basic and complex extraction methods to analyze over 200 million references in over 40 million articles from multilingual Wikipedia. We extracted information about the sources and unified them using special identifiers such as DOI, JSTOR, PMC, PMID, arXiv, ISBN, ISSN, OCLC and other. Additionally we used information about archive URL and included templates in the articles.

We proposed 10 models in order to assess popularity and reliability of websites, news magazines and other sources in Wikipedia. We also used DBpedia and Wikidata to automatically identify the alignment of the sources to specific field. Additionally, we analyzed the differences of popularity and reliability assessment of the sources between different periods. Moreover, we also conducted analysis of the growth leaders in each considered month. Results showed that depending on model and time some of the source can have different directions and power of changes (rise or fall). Next, we compared the similarity of rankings that used different models.

Some of extended results on reliability assessment of the sources in Wikipedia are placed in BestRef project [44].

In addition to what has already been described in the Section 10.2, in future work we plan to extend the popularity and reliability model. One of the directions is to take into account the position of the inserted reference in article and in list of the references. Next we plan to take into account features of the articles related to Wikipedia authors such as reputation or number of article watchers.

In this work we showed how it is possible to measure growth of the popularity and reliability of the sources based on differences in the Wikipedia content from several recent months. In our future research we plan to extend the time series to have more information about growth leaders in different years in each language version of Wikipedia.

Information about reliability of the sources can help to improve models for quality assessment of the Wikipedia articles. This can be especially useful to estimate sources of conflict statements between language versions of Wikipedia in articles related to the same subject. Additionally, one of the promising direction of the future work is to create methods for suggesting Wikipedia authors reliable sources for selected topics and statements in separate languages of Wikipedia.

Author Contributions: Conceptualization, W.L. and K.W.; methodology, W.L.; software, W.L.; validation, W.L. and K.W.; formal analysis, K.W. and W.A.; investigation, W.L.; resources, W.A.; data curation, W.L.; writing–original draft preparation, W.L.; writing–review and editing, K.W. and W.A.; visualization, W.L.; supervision, K.W. and W.A.; project administration, K.W. and W.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Position in Local Rankings

Table A1. Position in the local rankings of the most popular and reliable sources in different language versions of Wikipedia in February 2020 using **PR** model. Source: own calculations based on Wikimedia dumps using complex extraction of references. Extended version of the table is available on the web page: http://data.lewoniewski.info/sources/a1.

Source	Language Version of Wikipedia														
	ar	de	en	es	fa	fr	it	ja	nl	pl	pt	ru	sv	vi	zh
ad.nl	4169	166	633	11,663	6153	1086	5971	2737	3	1992	4003	7161	13,152	2142	12,739
adorocinema.com	4189	17,030	3731	1402	-	13,204	17,889	8990	16,592	141	2	15,003	20,774	5757	25,859
allocine.fr	2051	390	929	2138	901	2	565	1767	2323	1586	1488	963	517	4818	4491
almaany.com	3	23,568	5249	27,303	391	4592	18,098	21,354	-	-	10,374	7209	924	13,552	32,987
appledaily.com.tw	7260	24,734	3917	31,354	14,794	43,411	42,064	840	-	-	4103	31,323	-	426	2
cand.com.vn	26,768	80,003	47,951	-	-	-	-	-	-	-	-	75,342	-	3	18,821
deadline.com	7	2	1	1	2	1	2	8	11	5	1	2	20	5	1
dn.se	231	207	310	2174	2255	765	2011	3130	1223	1561	2165	1882	1	1109	1818
dwdl.de	1386	5	1359	19,652	8051	801	2716	26,042	5155	4579	27,221	32,027	5448	11,976	32,793
eiga.com	2719	7745	452	1609	3919	2000	3130	3	22,464	1528	926	2863	5463	174	33
elcinema.com	1	23,353	4628	38,243	1744	1585	25,524	40,045	12,266	14,817	35,232	12,767	7341	15,563	26,656
expressen.se	1392	557	300	1379	8263	389	487	6097	505	545	973	883	2	3011	1724
formulatv.com	112	1186	679	5	5705	323	202	59,424	22,695	5733	248	1171	25,332	24,837	32,378
hln.be	2052	3577	1817	17,379	15,411	1471	24,548	55,133	4	2069	5241	17,063	24,763	4085	4307
ibge.gov.br	-	18,761	13,284	2115	-	19,876	-	-	7030	-	4	4275	22,550	2902	38,937
imdb.com	2	4	4	4	4	7	13	44	12	4	8	6	4	15	13
infoescola.com	14,818	49,872	17,542	997	-	30,476	11,193	-	-	7107	5	44,201	24,945	5539	6575
irna.ir	1806	66,843	8072	20,057	1	38,803	66,342	42,350	17,815	-	16,456	21,773	-	11,503	17,543
kp.ru	3177	1809	874	6625	3459	2419	7793	3563	5480	634	13,005	4	5915	2236	1395
lenta.ru	352	325	462	930	1192	480	1254	785	2363	166	1342	1	1578	310	676
lesinrocks.com	1941	2308	1004	1600	1399	3	859	6069	2301	9497	3817	3074	9032	3804	2401
mobot.org	6862	125,005	4337	552	11,203	4969	5210	10,805	6734	2	1095	37,401	13,186	930	12,005
news.livedoor.com	2529	31,803	1628	2967	11,697	9632	13,447	5	-	24,057	10,329	6965	28,944	388	98
news.mynavi.jp	1522	5110	1394	12,368	4268	15,865	16,939	4	-	40,700	3880	11,560	7180	410	45
nikkei.com	3193	1096	694	5571	790	3854	1402	2	1977	4031	1524	3870	12,832	836	64
oricon.co.jp	226	360	60	167	686	121	347	1	2606	91	131	204	1115	22	3
regeringen.se	9566	12,561	4789	21,114	5065	68,510	-	64,855	17,468	33,056	4711	25,867	5	3017	45,773
repubblica.it	413	205	173	260	2403	136	3	1188	662	348	845	407	1221	1064	466
research.amnh.org	49,400	49,866	16,304	13,141	-	28,287	24,255	-	14	10,293	24,065	3317	-	2	24,727
rottentomatoes.com	16	10	5	9	18	11	19	50	44	6	9	7	109	30	14

Table A1. *Cont.*

Source	Language Version of Wikipedia														
	ar	**de**	**en**	**es**	**fa**	**fr**	**it**	**ja**	**nl**	**pl**	**pt**	**ru**	**sv**	**vi**	**zh**
scb.se	336	1248	777	3518	854	2800	1439	16,388	621	231	1759	1629	3	1234	1739
skijumping.pl	41,594	586	69,493	16,664	-	25,731	12,919	62,186	13,862	3	51,612	23,763	5186	-	42,126
taz.de	3959	3	1648	5397	785	692	3821	15,993	1918	996	13,190	1968	2268	577	3684
thefutoncritic.com	139	130	19	37	87	4	16	352	335	20	40	58	458	251	82
treccani.it	333	223	278	90	2344	59	1	2802	229	233	75	236	1786	871	2809
trouw.nl	9314	2869	2602	42,703	7579	1899	33,558	18,185	5	8491	13,875	22,557	24,774	16,870	27,600
tvbythenumbers.zap2it.com	45	22	3	15	33	6	5	19	119	143	7	5	165	49	12
tw.appledaily.com	37,437	23,163	10,245	53,429	-	58,799	-	1793	-	37,810	61,708	23,742	-	2001	5
universalis.fr	5	3273	3525	904	6465	8	1223	5180	2512	7534	871	2727	1012	11,754	18,729
variety.com	10	1	2	3	3	5	4	14	13	7	3	3	19	4	4
vnexpress.net	13,310	18,184	6504	58,271	7212	9972	39,942	9639	19,417	30,018	28,707	13,486	12,178	1	9857
volkskrant.nl	2766	918	949	6873	3345	1781	3775	10,507	2	12,197	5107	2687	16,051	4644	15,292
web.archive.org	4	36	35	2	12	18	24	12	1	1	17	18	14	10	57
who.int	11	13	67	13	5	26	31	32	29	38	28	63	28	6	10

Table A2. Position in local rankings of periodical sources in different language versions of Wikipedia in February 2020 using **PR** model. Source: own work based on Wikimedia dumps using complex extraction of references using complex extraction of references with semantic databases (Wikidata, DBpedia) to identify type of the source. Extended version of the table is available on the web page: http://data.lewoniewski.info/sources/a2.

Source	Position in Local Rankings in Language Versions of Wikipedia														
	ar	**de**	**en**	**es**	**fa**	**fr**	**it**	**ja**	**nl**	**pl**	**pt**	**ru**	**sv**	**vi**	**zh**
20 minutos	176	186	189	2	95	87	81	265	64	119	46	252	333	232	262
Aftonbladet	1657	1504	970	1484	546	132	1117	1013	981	708	1111	1270	10	-	1369
Al-Ittihad	10	1530	2972	2731	537	-	-	1397	-	-	-	1672	-	-	1290
Algemeen Dagblad	387	43	191	795	373	182	438	143	2	184	300	538	714	202	753
Aliqtisadi	2	-	2022	-	669	-	2338	1138	-	-	-	1096	-	-	-
Apple Daily	562	1233	644	1406	807	1768	1561	73	-	1525	308	1275	-	56	1
Auto, Motor und Sport	1152	4	535	373	-	727	376	221	275	136	-	487	585	428	-
China Press	2227	-	1420	2356	-	1241	-	431	-	-	1567	2025	-	254	8
DWDL.de	162	3	361	1073	471	145	270	764	362	315	1119	1296	386	767	1430
Dagens Industri	1336	1133	949	1682	-	292	1572	589	-	1114	623	2531	2	-	1376
De Gelderlander	923	397	1026	1774	455	628	1370	1030	10	459	1104	1014	-	824	637
De Morgen	682	212	508	577	157	210	593	412	6	254	440	830	473	355	480
De Stentor	1380	418	1428	-	-	1223	2055	1947	9	1333	-	1898	795	641	1696
De Volkskrant	293	145	283	575	221	272	330	403	1	669	355	299	808	373	847
Die Tageszeitung	374	2	414	496	80	130	332	539	206	114	732	231	197	66	298
Donya-e-Eqtesad	1272	-	2665	2805	2	-	2193	-	-	-	-	2833	-	-	1378
El Confidencial	243	226	219	3	281	57	98	485	253	243	83	235	321	264	190
El País	217	224	400	6	205	146	263	404	480	86	84	260	401	309	562
Ennahar newspaper	5	-	2248	-	727	1042	-	-	-	-	-	-	-	-	-
Entertainment Weekly	8	6	2	4	6	7	7	13	13	8	4	6	11	4	5
Exame	779	1474	683	610	554	1179	466	797	949	1115	3	1917	641	85	759
Expert	192	1085	936	1177	729	1385	1501	542	749	490	2053	10	-	589	926
Express Gazeta	882	941	1045	1301	302	1502	1193	907	790	621	1734	8	-	609	824
Famitsu	2004	1810	503	558	1019	755	703	10	-	1599	605	693	1093	624	38
Finanztest	229	7	1404	1836	213	1704	901	565	174	1006	-	1329	909	-	316
Flight International	32	10	23	73	28	44	59	19	70	11	49	87	101	37	25
Fokus	501	1538	1380	1209	959	944	2054	961	-	1316	761	1315	6	-	-
Folha de S. Paulo	119	1082	652	304	621	958	306	1634	812	748	7	1697	-	490	834
Fortune	29	32	16	45	25	25	54	36	66	41	29	65	103	8	23
Gazeta do Povo	1429	745	1066	385	-	1044	1011	1257	-	-	9	2115	654	1123	-

Table A2. *Cont.*

Source	Position in Local Rankings in Language Versions of Wikipedia														
	ar	**de**	**en**	**es**	**fa**	**fr**	**it**	**ja**	**nl**	**pl**	**pt**	**ru**	**sv**	**vi**	**zh**
Helsingborgs Dagblad	505	804	857	968	391	717	1049	588	853	766	279	390	3	154	694
Het Laatste Nieuws	214	399	430	999	836	229	1096	762	3	188	361	900	1162	331	341
Het Parool	1149	337	550	586	427	492	933	386	7	1663	630	1740	1116	459	538
Huffington Post France	569	535	599	405	334	6	308	601	220	240	392	451	575	253	1759
ISTOÉ	851	997	1130	668	-	1217	919	833	1125	732	8	959	495	680	1106
Il Fatto Quotidiano	313	126	230	211	508	147	4	682	765	226	353	346	636	475	663
Il Post	540	207	569	332	693	218	3	181	536	263	299	372	436	435	440
Jeune Afrique	39	200	342	210	212	4	224	463	229	425	215	364	313	276	413
Komsomolskaya Pravda	226	187	177	418	155	120	273	131	352	52	397	2	350	133	140
la Repubblica	63	15	45	56	82	29	1	65	110	45	91	73	137	59	62
La Tercera	269	487	417	7	609	499	169	695	339	1311	172	511	745	379	810
Le Figaro	511	285	563	321	493	5	279	717	845	1134	387	470	419	1009	398
Le Monde	159	244	306	300	159	3	246	499	248	567	325	424	639	292	351
Lenta.ru	60	67	142	162	92	105	166	69	224	24	139	1	157	43	98
Les Inrockuptibles	211	287	293	233	119	1	129	264	222	570	283	322	547	322	228
NRC Next	843	344	539	1113	248	687	884	1049	5	707	674	837	230	445	173
Nauka i Zhizn	-	2536	610	421	371	1431	506	-	1040	1635	289	7	-	1205	1810
Nguoi Viet Daily News	1126	-	1851	-	-	1064	-	-	-	-	-	-	-	6	858
Nihon Keizai Shimbun	322	169	206	503	81	423	177	1	208	279	157	368	698	90	14
Nikkei Business	2409	1314	898	2079	-	2747	1271	8	-	-	1410	2597	-	750	306
Nishinippon Shimbun	-	1292	2092	1248	-	3266	1786	7	-	-	-	1576	-	1160	115
O Estado de São Paulo	897	1586	1020	590	-	611	1144	1200	-	1352	5	1385	728	242	829
PC Gamer	51	51	20	30	10	31	55	51	90	17	26	12	53	14	12
PC Games	1785	8	635	936	-	1023	908	1685	563	306	849	280	441	425	534
Panorama	565	726	506	534	885	341	10	1256	-	336	734	607	-	351	425
People	25	5	6	12	13	8	13	26	14	6	16	4	19	36	18
Pitchfork	117	28	7	20	40	15	25	37	36	26	20	28	25	26	55
Populär Historia	1299	671	1844	2420	438	2514	2302	1113	-	998	186	1218	7	-	1096
Rolling Stone	76	21	10	13	16	20	15	27	11	14	21	25	28	30	44
Rolling Stone Brasil	1656	2310	709	695	1063	853	637	949	756	1157	10	324	1018	747	905

Table A2. *Cont.*

Source	Position in Local Rankings in Language Versions of Wikipedia														
	ar	**de**	**en**	**es**	**fa**	**fr**	**it**	**ja**	**nl**	**pl**	**pt**	**ru**	**sv**	**vi**	**zh**
Sai Gon Giai Phong	714	2509	2084	-	443	1840	2217	1680	-	1797	830	1535	-	3	712
Sport Express	367	295	344	490	399	313	344	179	444	100	458	5	626	382	514
Superinteressante	1799	3132	804	991	-	754	1720	608	-	1468	6	-	-	446	-
Svenska Dagbladet	409	1997	1495	1596	-	2789	795	1158	-	-	89	1237	9	-	414
Sydsvenskan	495	385	566	818	70	594	700	430	547	514	305	598	1	782	895
TV Guide	61	44	17	54	63	46	50	103	59	3	28	39	184	38	54
TV Sorrisi e Canzoni	153	386	618	69	712	453	6	968	180	591	512	632	335	730	1312
TechCrunch	7	17	11	9	14	14	27	9	29	16	12	23	24	5	11
Teknikens Värld	-	408	1366	-	181	1469	1357	-	1113	1607	-	1176	5	-	545
The Atlantic	21	25	12	25	7	24	46	32	31	37	24	42	33	20	35
The Daily Telegraph	14	12	8	18	8	16	14	17	18	9	15	16	17	9	17
The Indian Express	28	84	5	135	9	92	153	90	156	61	87	90	147	40	57
The New York Times	15	27	14	16	18	22	38	23	34	38	23	38	21	13	3
The Washington Post	3	13	3	19	4	9	18	24	28	25	22	29	18	12	20
Time	4	11	9	10	3	11	20	18	22	10	14	15	13	7	10
Tokyo Sports	731	1981	270	404	1020	599	527	2	-	832	769	1006	-	181	19
Trouw	704	334	543	1687	444	279	1346	587	4	526	752	1053	1163	1033	1276
USA Today	6	9	4	8	5	10	19	14	15	12	11	20	8	10	7
Variety	1	1	1	1	1	2	2	4	8	1	1	3	4	2	2
Veja	356	558	442	199	479	378	816	866	969	550	2	619	621	394	755
VnExpress	920	1018	977	2021	429	745	472	371	982	1270	1171	768	671	1	633
Vokrug sveta	1906	1183	1378	2121	-	867	1055	901	865	220	-	9	-	372	687
Weekly Playboy	1159	-	1581	549	-	2036	1312	5	-	1344	-	1499	-	289	31
Wired	9	20	13	14	11	18	9	6	37	13	18	21	29	11	15
World Journal	-	-	714	908	-	-	1307	190	-	-	-	-	-	1096	4
Wprost	741	632	945	855	908	1281	1278	665	980	2	930	544	1004	439	795
Yomiuri Shimbun	273	1010	372	911	563	592	565	3	501	1368	828	1055	-	367	43
¡Hola!	181	204	185	5	289	128	91	229	789	110	57	207	124	273	331

References

1. Wikipedia Meta-Wiki. List of Wikipedias. Available online: https://meta.wikimedia.org/wiki/List_of_Wikipedias (accessed on 30 March 2020).
2. English Wikipedia. Reliable Sources. Available online: https://en.wikipedia.org/wiki/Wikipedia:Reliable_sources (accessed on 30 March 2020).
3. Internet Live Stats. Total Number of Websites. Available online: https://www.internetlivestats.com/total-number-of-websites/ (accessed on 30 March 2020).
4. Eysenbach, G.; Powell, J.; Kuss, O.; Sa, E.R. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *JAMA* **2002**, *287*, 2691–2700. [CrossRef] [PubMed]
5. Price, R.; Shanks, G. A semiotic information quality framework: development and comparative analysis. In *Enacting Research Methods in Information Systems*; Springer: Berlin, Germany, 2016; pp. 219–250.
6. Xu, J.; Benbasat, I.; Cenfetelli, R.T. Integrating service quality with system and information quality: An empirical test in the e-service context. *MIS Q.* **2013**, *37*, 777–794. [CrossRef]
7. Nielsen, F.Å. Scientific citations in Wikipedia. *arXiv* **2007**, arXiv:0705.2106.
8. Lewoniewski, W.; Węcel, K.; Abramowicz, W. Analysis of references across Wikipedia languages. In Proceedings of the International Conference on Information and Software Technologies, Druskininkai, Lithuania, 12–14 October 2017; pp. 561–573.
9. Characterizing Wikipedia Citation Usage. Analyzing Reading Sessions. Available online: https://meta.wikimedia.org/wiki/Research:Characterizing_Wikipedia_Citation_Usage/Analyzing_Reading_Sessions (accessed on 29 February 2020).
10. Jemielniak, D.; Masukume, G.; Wilamowski, M. The most influential medical journals according to Wikipedia: quantitative analysis. *J. Med. Internet Res.* **2019**, *21*, e11429. [CrossRef] [PubMed]
11. Stvilia, B.; Twidale, M.B.; Smith, L.C.; Gasser, L. Assessing information quality of a community-based encyclopedia. *Proc. ICIQ* **2005**, pp. 442–454.
12. Blumenstock, J.E. Size matters: Word count as a measure of quality on Wikipedia. In Proceedings of the 17th International Conference on World Wide Web, Beijing, China, 21–25 April 2008; pp. 1095–1096.
13. Lucassen, T.; Schraagen, J.M. Trust in wikipedia: How users trust information from an unknown source. In Proceedings of the 4th Workshop on Information Credibility, Raleigh, NC, USA, 27 April 2010; pp. 19–26.
14. Yaari, E.; Baruchson-Arbib, S.; Bar-Ilan, J. Information quality assessment of community generated content: A user study of Wikipedia. *J. Inf. Sci.* **2011**, *37*, 487–498. [CrossRef]
15. Conti, R.; Marzini, E.; Spognardi, A.; Matteucci, I.; Mori, P.; Petrocchi, M. Maturity assessment of Wikipedia medical articles. In Proceedings of the 2014 IEEE 27th International Symposium on Computer-Based Medical Systems (CBMS), New York, NY, USA, 27–29 May 2014; pp. 281–286.
16. Piccardi, T.; Redi, M.; Colavizza, G.; West, R. Quantifying Engagement with Citations on Wikipedia. *arXiv* **2020**, arXiv:2001.08614.
17. Nielsen, F.Å.; Mietchen, D.; Willighagen, E. Scholia, scientometrics and wikidata. In Proceedings of the European Semantic Web Conference, Portorož, Slovenia, 28 May–1 June 2017; pp. 237–259.
18. Teplitskiy, M.; Lu, G.; Duede, E. Amplifying the impact of open access: Wikipedia and the diffusion of science. *J. Assoc. Inf. Sci. Technol.* **2017**, *68*, 2116–2127. [CrossRef]
19. Fetahu, B.; Markert, K.; Nejdl, W.; Anand, A. Finding news citations for wikipedia. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, Indianapolis, IN, USA, 24–28 October 2016; pp. 337–346.
20. Ferschke, O.; Gurevych, I.; Rittberger, M. FlawFinder: A Modular System for Predicting Quality Flaws in Wikipedia. In Proceedings of the CLEF (Online Working Notes/Labs/Workshop), Rome, Italy, 17–20 September 2012; pp. 1–10.
21. Flekova, L.; Ferschke, O.; Gurevych, I. What makes a good biography?: multidimensional quality analysis based on wikipedia article feedback data. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea, 7–11 April 2014; pp. 855–866.
22. Shen, A.; Qi, J.; Baldwin, T. A Hybrid Model for Quality Assessment of Wikipedia Articles. In Proceedings of the Australasian Language Technology Association Workshop, Brisbane, Australia, 6–8 December 2017; pp. 43–52.

23. di Sciascio, C.; Strohmaier, D.; Errecalde, M.; Veas, E. WikiLyzer: interactive information quality assessment in Wikipedia. In Proceedings of the 22nd International Conference on Intelligent User Interfaces, Limassol, Cyprus, 13–16 March 2017; pp. 377–388.
24. Dang, Q.V.; Ignat, C.L. Measuring Quality of Collaboratively Edited Documents: The Case of Wikipedia. In Proceedings of the 2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC), Pittsburgh, PA, USA, 31 October–3 November 2016; pp. 266–275.
25. Lewoniewski, W.; Węcel, K.; Abramowicz, W. Relative Quality and Popularity Evaluation of Multilingual Wikipedia Articles. *Informatics* **2017**, *4*, 43. [CrossRef]
26. Lewoniewski, W.; Węcel, K.; Abramowicz, W. Multilingual Ranking of Wikipedia Articles with Quality and Popularity Assessment in Different Topics. *Computers* **2019**, *8*, 60. [CrossRef]
27. Warncke-wang, M.; Cosley, D.; Riedl, J. Tell Me More: An Actionable Quality Model for Wikipedia. In Proceedings of the WikiSym 2013, Hong Kong, China, 5–7 August 2013; pp. 1–10.
28. Lih, A. Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource. In Proceedings of the 5th International Symposium on Online Journalism, Austin, TX, 16-17 April, 2004; p. 31.
29. Liu, J.; Ram, S. Using big data and network analysis to understand Wikipedia article quality. *Data Knowl. Eng.* **2018**, *115*, 80–93. [CrossRef]
30. Wilkinson, D.M.; Huberman, B.A. Cooperation and quality in wikipedia. In Proceedings of the 2007 International Symposium on Wikis WikiSym 07, Montreal, QC, Canada, 21–23 October 2007; pp. 157–164. [CrossRef]
31. Kane, G.C. A multimethod study of information quality in wiki collaboration. *ACM Trans. Manag. Inf. Syst. (TMIS)* **2011**, *2*, 4. [CrossRef]
32. WikiTop. Wikipedians Top. Available online: http://wikitop.org/ (accessed on 30 March 2020).
33. Lewoniewski, W. The Method of Comparing and Enriching Information in Multlingual Wikis Based on the Analysis of Their Quality. Ph.D. Thesis, Poznań University of Economics and Business, Poznań, Poland, 2018.
34. Lerner, J.; Lomi, A. Knowledge categorization affects popularity and quality of Wikipedia articles. *PLoS ONE* **2018**, *13*, e0190674. [CrossRef] [PubMed]
35. Wikimedia Downloads. English Wikipedia Latest Database Backup Dumps. Available online: https://dumps.wikimedia.org/enwiki/latest/ (accessed on 30 March 2020).
36. English Wikipedia. 2019–2020 Coronavirus Pandemic. Available online: https://en.wikipedia.org/wiki/2019%E2%80%9320_coronavirus_pandemic (accessed on 30 March 2020).
37. Vrandečić, D.; Krötzsch, M. Wikidata: A free collaborative knowledgebase. *Commun. ACM* **2014**, *57*, 78–85. [CrossRef]
38. Wikidata. Available online: https://www.wikidata.org/wiki/Wikidata:Main_Page (accessed on 23 April 2020).
39. Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. DBpedia: A Nucleus for a Web of Open Data. In *The Semantic Web;* Aberer, K., Choi, K.S., Noy, N., Allemang, D., Lee, K.I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., et al., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 722–735.
40. DBpedia. Available online: https://wiki.dbpedia.org/ (accessed on 23 April 2020).
41. Frey, J.; Hofer, M.; Obraczka, D.; Lehmann, J.; Hellmann, S. DBpedia FlexiFusion the Best of Wikipedia> Wikidata> Your Data. In Proceedings of the International Semantic Web Conference, Auckland, New Zealand, 26–30 October 2019; pp. 96–112.
42. GFS Data Browser. Available online: https://global.dbpedia.org (accessed on 23 April 2020).
43. English Wikipedia. Perennial Sources. Available online: https://en.wikipedia.org/wiki/Wikipedia:Reliable_sources/Perennial_sources (accessed on 30 March 2020).
44. BestRef. Popular and Reliable Sources of Wikipedia. Available online: https://bestref.net (accessed on 30 March 2020).