Peer-reviewed version available at *Information* **2020**, *11*; doi:10.3390/info11050263

Article

Modeling Popularity and Reliability of Sources in Multilingual Wikipedia

Włodzimierz Lewoniewski [1], Krzysztof Węcel [1] and Witold Abramowicz [1]

[1] Department of Information Systems, Poznań University of Economics and Business, Poland; {wlodzimierz.lewoniewski,krzysztof.wecel,witold.abramowicz}@ue.poznan.pl

Abstract: One of the most important factors impacting quality of content in Wikipedia is presence of credible 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 nearly 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 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 versions 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 version 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 own rules and criteria of reliability, as well as 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 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 met 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 of 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.” Therefore, we decide 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. We 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.

2. Recent Work

Due to the fact that source reliability is an 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 references count 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 [7,8]. Based on the references users can assess the trustworthiness of Wikipedia articles, therefore we consider the source of information is an important factor [9].

Often references contain 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 [10,11].

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 [12]. 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 [13]. There are also works, which showed that a lot of citations in Wikipedia articles refers to scientific publications [12,14], especially if they are open-access [15], wherein Wikipedia authors prefer to put recently published journal articles as a source [16]. 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 [17].

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 [11,18–24]. 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 [25].

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 [26,27]. Therefore, we can use the number of unique authors as one of the measures of quality of Wikipedia articles [24,28,29].

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 [30]. Therefore, popularity can play an important role not only for quality estimation of information in specific language version of Wikipedia [31] but also for checking reliability of the sources in it. Larger number of readers of Wikipedia article may allow for more rapid changes in incorrect or outdated information [24]. Popularity of an article can be measured based on the number of visits [32].

Taking into account different studies related to references 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 time 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 [33]), 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 of source usage.
2. **P** model – based on cumulative pageviews of the article in which source appears.
3. **PR** model – based on cumulative pageviews of the article in which source appears divided by number of the references in this article.
4. **PL** model – based on cumulative pageviews of the article in which source appears divided by article length.
5. **Pm** model – based on daily pageviews median of the article in which source appears.
6. **PmR** model – based on daily pageviews median of the article in which source appears divided by number of the references in this article.
7. **PmL** model – based on daily pageviews median of the article in which source appears divided by article length.
8. **A** model – based on number of authors of the article in which source appears.
9. **AR** model – based on number of authors of the article in which source appears divided by number of the references in this article.
10. **AL** model – based on number of authors of the article in which source appears divided by article length.

Frequency of source usage in **F** model means how many references contain the analyzed domain in URL. 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 [31]. We can expect that articles read by many people are more likely to have verified and credible 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 [24] we decided to measure pageviews also as daily pageviews median (abbr. "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 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, 6, 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 March 1, 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

Often references have 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.

Lang.	Number of articles				Number of URLs	
	all	with >=1 URL	with >= 10 URLs	with >=100 URLs	all	unique
ar	1 031 740	917 809	305 118	4 369	9 443 788	7 599 390
az	156 442	109 743	20 299	237	674 212	512 465
be	185 753	150 116	21 067	299	1 142 005	958 165
bg	260 081	211 031	27 806	185	1 174 324	1 030 715
ca	638 664	600 711	336 302	1 770	8 111 104	7 124 746
cs	447 120	377 647	69 821	1 220	2 769 415	2 438 870
da	257 321	211 415	51 689	488	1 711 677	1 605 379
de	2 403 683	1 990 310	528 524	7 849	17 646 882	15 632 584
el	174 589	151 008	43 664	891	1 479 933	1 254 224
en	6 029 201	5 500 527	1 963 703	60 384	69 554 575	56 030 670
eo	275 674	223 652	21 028	85	1 016 902	928 935
es	1 528 811	1 395 107	484 650	5 521	13 935 332	11 872 312
et	206 430	136 651	8 344	146	526 292	466 916
eu	349 176	331 836	97 469	104	2 692 639	2 177 612
fa	712 216	656 161	52 779	1 030	2 779 293	2 232 907
fi	479 830	405 372	61 387	545	2 446 538	1 889 702
fr	2 185 885	1 830 876	593 874	7 327	17 918 673	15 313 234
gl	161 860	127 395	52 159	595	1 483 541	1 315 467
he	261 209	213 989	76 274	347	2 152 942	1 987 360
hi	140 327	97 706	10 102	370	563 963	379 306
hr	198 670	137 949	10 796	155	587 017	449 783
hu	465 509	411 072	97 289	1 179	3 231 880	2 796 234
hy	264 676	219 045	50 681	1 218	2 073 940	1 534 220
id	524 100	409 937	53 085	1 267	2 496 158	2 158 397
it	1 586 855	1 374 018	403 171	3 194	11 889 377	10 141 992
ja	1 192 596	890 138	205 264	4 210	7 449 642	6 309 830
ka	135 333	102 910	10 508	239	533 019	420 322
kk	230 376	137 333	6 536	54	736 786	591 481
ko	486 067	318 190	63 425	1 110	2 197 777	1 990 960
la	132 258	106 887	3 592	22	347 131	287 532
lt	196 606	136 982	4 238	27	390 006	331 424
ms	335 222	191 206	18 288	431	868 166	716 712
nl	1 999 092	1 626 602	31 700	1 460	4 303 813	3 295 204
nn	151 857	126 229	16 642	73	624 568	561 283
no	529 426	466 557	132 817	672	3 812 791	3 410 905
pl	1 387 164	1 177 588	159 956	2 334	6 962 407	5 673 526
pt	1 022 524	925 771	186 889	4 454	7 836 416	6 583 420
ro	404 748	352 338	80 111	970	2 742 321	2 375 095
ru	1 602 761	1 333 264	527 323	8 184	16 116 795	12 370 583
sh	451 298	383 945	223 652	292	4 464 569	1 118 996
simple	155 887	103 886	10 990	264	548 488	480 654
sk	232 551	176 188	10 893	268	823 474	681 781
sl	167 119	135 614	21 910	219	786 235	710 113
sr	630 870	552 584	53 185	761	3 502 213	1 959 054
sv	3 740 411	3 590 906	798 561	2 356	21 372 068	11 686 205
ta	132 424	105 186	10 658	228	569 482	401 066
th	135 627	93 945	16 965	726	758 451	667 308
tr	343 216	257 976	40 305	1 306	1 762 805	1 495 178
uk	994 030	859 711	185 470	2 476	6 973 455	5 195 088
ur	154 282	120 189	5 229	191	403 727	354 010
uz	133 774	92 369	964	27	299 080	265 877
vi	1 241 487	1 178 177	46 835	1 580	3 604 033	2 846 271
vo	124 189	93 924	9	-	104 201	103 660
zh	1 099 744	862 260	175 496	4 873	6 757 646	5 779 801
zh-min-nan	267 615	192 933	519	1	353 098	274 056

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.

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 one 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).

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.g. 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 more complex extraction based on source code of each Wikipedia article. This method 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 [24]. 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. Among the most commonly used templates in this Wikipedia language versions are: 'Cite web', 'Cite news', 'Cite book', 'Cite journal', 'NHLE' (National Heritage List for England), 'Citation', 'Webarchive', 'ISBN', 'In lang', 'Dead link', 'Harvnb' (Harvard citation no brackets), 'Cite magazine'.

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

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.

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 "2019–20 coronavirus pandemic" that was added using template "2019–20 coronavirus pandemic data". In our approach we include such references in the analysis when such templates appear in the Wikipedia articles.

Some of the most popular templates allows to add identifiers to the source such as DOI, JSTOR, PMC, PMID, arXiv, ISBN, ISSN, OCLC and other. 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.

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.

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

Lang.	DOI		ISBN		ISSN		PMID		PMC	
	all	unique	all	unique	all	unique	all	unique	all	unique
ar	129 274	87 067	161 332	77 217	23 644	7 634	83 075	58 027	18 350	12 777
az	2 148	1 261	20 708	8 289	842	241	524	379	127	106
be	2 442	1 544	45 271	5 893	941	256	1 120	735	165	111
bg	8 328	5 794	48 387	13 921	1 277	488	6 023	3 744	989	700
ca	47 665	33 602	177 475	69 265	22 780	6 478	27 384	21 589	7 788	6 334
cs	26 331	15 872	175 552	33 259	28 775	4 657	12 268	7 794	1 316	925
da	7 399	4 601	28 475	12 474	1 392	520	4 873	2 856	892	556
de	158 399	82 168	890 727	199 949	77 065	13 250	18 893	12 821	14 660	9 284
el	22 586	14 352	62 201	26 531	4 635	1 505	12 315	7 761	2 510	1 649
en	2 066 518	908 776	3 585 981	803 390	513 124	37 894	986 033	476 603	343 970	156 515
eo	4 804	3 175	18 121	9 458	682	327	1 922	1 249	565	340
es	130 601	75 796	432 030	152 357	71 640	12 528	65 335	39 281	14 779	8 441
et	7 481	3 269	16 419	5 117	534	171	4 809	2 063	1 134	509
eu	7 089	5 081	16 928	9 371	6 778	2 194	1 503	1 183	5 905	3 621
fa	26 876	18 547	43 088	20 402	5 160	2 134	17 146	11 351	4 490	2 997
fi	10 151	5 394	177 952	25 085	7 991	1 954	5 182	2 936	370	276
fr	156 864	71 915	816 380	191 909	120 452	15 291	39 749	25 937	4 272	2 296
gl	44 742	30 259	49 590	19 831	5 433	1 795	34 897	24 793	10 272	7 090
he	8 598	7 747	12 525	9 553	1 589	412	3 883	3 632	590	546
hi	10 854	6 964	26 204	12 073	1 003	435	9 146	6 204	1 427	1 013
hr	5 565	3 316	19 982	7 917	904	291	4 876	2 727	726	471
hu	21 915	14 417	64 179	20 329	4 220	1 524	12 106	8 128	2 075	1 454
hy	36 496	22 056	51 252	24 089	6 442	2 296	32 273	19 077	7 626	4 566
id	35 011	21 094	108 227	42 208	8 711	2 305	16 553	10 698	3 920	2 695
it	90 425	51 156	425 878	95 326	14 108	3 428	45 394	32 052	7 030	4 642
ja	115 036	57 437	463 509	86 153	25 281	5 303	42 622	27 090	11 167	6 586
ka	3 589	2 539	12 750	5 974	1 225	337	1 393	1 083	297	238
kk	375	312	55 903	1 323	67	39	210	178	78	63
ko	40 349	20 692	55 431	22 932	6 249	1 618	14 483	9 364	3 885	2 451
la	700	521	2 870	1 860	49	36	294	245	74	55
lt	1 456	1 083	10 743	3 572	315	137	940	655	187	144
ms	11 252	7 628	25 413	13 858	1 726	663	5 678	3 931	1 416	945
nl	12 653	8 529	45 406	16 246	1 767	817	7 496	5 339	1 470	1 036
nn	3 387	1 777	17 645	6 344	591	179	769	478	165	108
no	11 638	6 416	66 958	25 602	7 040	1 388	4 882	3 157	734	510
pl	130 714	41 832	497 638	60 329	70 513	8 869	48 949	27 897	6 904	3 694
pt	84 360	45 449	257 560	80 044	34 735	6 953	33 744	20 767	7 099	4 451
ro	18 466	11 504	54 371	21 353	2 926	1 012	10 728	6 475	2 250	1 504
ru	122 216	59 601	452 487	112 558	61 063	10 452	36 920	22 826	7 095	4 349
sh	53 715	12 715	40 709	10 500	2 836	551	28 975	21 897	3 220	2 219
simple	7 729	5 336	25 113	13 276	2 100	610	4 265	2 953	908	668
sk	3 154	2 231	39 734	8 805	7 044	1 130	735	569	127	106
sl	12 782	7 930	42 400	11 977	1 521	668	9 519	5 993	1 689	1 104
sr	66 733	21 939	72 201	24 295	5 297	1 870	35 498	26 272	5 002	3 324
sv	863 270	8 921	143 517	26 925	11 474	2 395	6 599	3 688	1 335	816
ta	19 643	13 982	27 372	15 364	1 688	747	11 118	8 153	2 083	1 384
th	26 407	16 144	30 447	14 406	2 840	927	18 952	11 727	4 248	2 681
tr	18 302	10 993	47 227	21 099	3 211	998	9 308	5 995	1 719	1 195
uk	254 338	24 397	107 702	35 669	53 199	3 263	55 205	10 129	3 222	2 216
ur	1 478	896	7 943	4 598	362	138	546	379	157	106
uz	144	126	868	541	25	19	26	24	11	10
vi	70 339	39 515	122 919	42 363	10 130	2 680	32 651	21 664	8 645	5 652
vo	-	-	87	77	-	-	-	-	-	-
zh	107 698	59 098	336 538	89 457	25 109	6 268	48 012	29 761	11 067	6 878
zh-min-nan	289	162	610	256	20	11	62	51	20	15

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.

Lang.	arXiv		Bibcode		JSTOR		LCCN		OCLC	
	all	unique	all	unique	all	unique	all	unique	all	unique
ar	8 585	3 014	21 033	9 914	4 747	3 569	413	280	7 656	4 956
az	144	72	793	389	398	109	50	26	414	161
be	253	129	547	318	52	38	4	3	41	36
bg	404	309	1 392	1 087	272	215	26	20	643	297
ca	1 722	905	5 487	3 320	1 430	931	141	86	2 609	1 277
cs	1 434	579	3 803	1 713	206	138	24	17	6 679	2 500
da	160	84	750	537	235	164	82	36	1 029	352
de	6 430	3 318	7 586	3 591	3 789	2 060	266	116	4 633	2 516
el	1 829	872	5 257	2 959	939	509	124	53	1 293	593
en	154 122	28 647	388 026	116 884	147 695	66 921	16 808	4 360	229 093	72 283
eo	39	20	241	179	356	253	22	21	199	155
es	2 888	1 646	12 019	7 185	4 151	2 516	511	243	26 725	12 040
et	320	132	1 355	597	134	68	9	5	147	75
eu	185	51	380	159	50	43	4	3	170	115
fa	898	533	3 195	2 129	669	510	63	38	1 538	833
fi	110	89	460	345	164	104	38	28	133	100
fr	11 122	2 942	23 110	6 995	4 353	2 502	6 934	2 565	58 798	20 787
gl	831	340	3 893	2 323	1 258	786	560	186	2 313	1 143
he	70	68	344	315	244	225	9	9	1 593	1 215
hi	1 061	272	2 185	772	218	153	55	35	562	341
hr	166	124	690	522	114	79	4	4	396	159
hu	357	243	1 597	1 159	582	436	59	43	1 235	464
hy	428	253	3 648	1 681	652	507	68	35	1 742	788
id	2 288	813	7 581	3 600	1 871	1 046	213	108	5 839	2 170
it	2 845	1 291	5 861	3 610	1 918	1 138	2 445	684	20 123	7 204
ja	11 193	3 045	32 914	9 351	2 390	1 439	581	204	8 977	3 477
ka	425	269	1 143	802	158	114	66	50	435	241
kk	36	20	55	50	17	14	-	-	20	16
ko	7 617	2 562	15 107	5 505	1 265	831	253	99	1 367	581
la	4	4	52	45	47	34	6	6	44	30
lt	122	79	196	147	47	38	1	1	105	49
ms	657	374	2 379	1 567	462	340	54	41	999	541
nl	35	28	317	261	163	123	22	5	331	270
nn	743	168	1 632	560	135	94	13	3	196	117
no	919	223	2 734	1 112	384	242	42	25	1 465	593
pl	2 448	840	5 325	2 163	1 223	610	141	46	26 087	7 147
pt	4 260	1 666	19 581	5 998	2 805	1 793	322	167	11 096	4 791
ro	1 181	495	4 004	2 269	636	447	175	84	1 510	710
ru	11 798	2 998	25 274	7 545	1 835	1 061	331	115	3 292	1 449
sh	171	91	1 101	720	272	205	35	17	2 981	623
simple	543	264	1 221	824	227	177	38	26	1 011	391
sk	198	131	398	291	16	13	10	3	334	160
sl	662	185	1 454	667	287	255	16	15	453	306
sr	628	410	2 963	2 059	899	672	94	68	3 456	1 717
sv	1 042	391	3 087	1 251	293	219	179	22	5 029	1 623
ta	697	306	2 624	1 662	542	368	84	43	849	453
th	1 052	339	2 853	1 504	479	322	32	23	894	471
tr	2 148	769	5 267	2 387	599	357	107	59	973	514
uk	4 254	1 741	14 450	5 199	835	596	167	69	2 699	1 373
ur	93	33	208	127	137	102	9	7	231	144
uz	24	20	93	79	6	5	16	4	14	11
vi	7 781	2 639	18 635	7 859	2 280	1 364	292	188	5 245	1 616
vo	-	-	-	-	-	-	-	-	-	-
zh	11 370	3 463	26 908	10 365	2 508	1 533	392	233	9 927	3 160
zh-min-nan	1	1	82	43	8	6	-	-	9	9

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..

.

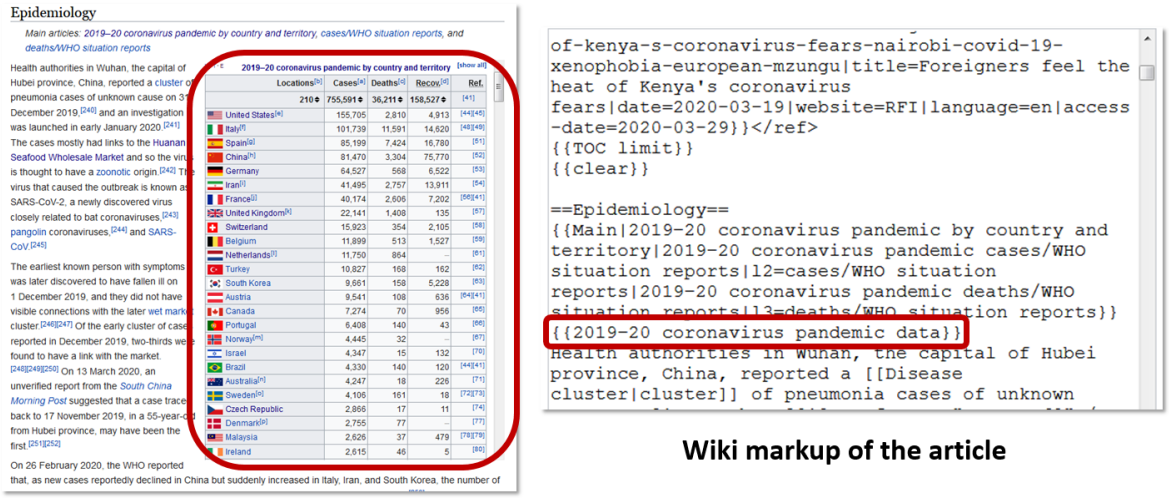

Figure 3. Table table with references in the Wikipedia article ”2019–20 coronavirus pandemic” that was added using template ”2019–20 coronavirus pandemic data”. Source [34]

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/...*

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

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 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 parameter depending on the archive service: „wayback", „archive-is", „webciteID" and other. So 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 other corresponding name and additional synonymous names. For example in English Wikipedia „Harvard citation", "Harv" and "Harvsp" means 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.

Analysis of the numbers of the references extracted by complex extraction showed other statistics comparing to basic extraction of the external links described in subsection 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 has each article in average. 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%.

Lang.	Number of articles				Number of references	
	all	with >= 1 ref.	with >= 10 refs.	with >= 100 refs.	all	unique
ar	1 031 740	817 240	58 060	2 551	3 583 772	2 136 477
az	156 442	77 162	6 450	427	426 141	209 781
be	185 753	90 340	5 822	253	346 788	162 999
bg	260 081	152 549	12 993	304	695 503	397 218
ca	638 664	418 730	53 386	1 300	2 586 974	1 322 428
cs	447 120	229 689	45 777	1 258	1 759 910	911 034
da	257 321	99 140	13 112	596	609 648	395 395
de	2 403 683	1 350 469	276 204	6 214	10 343 100	6 150 128
el	174 589	100 479	23 932	981	963 961	588 558
en	6 029 201	4 720 740	1 341 402	63 816	57 583 024	28 874 374
eo	275 674	54 836	4 089	149	229 984	152 861
es	1 528 811	1 071 838	226 781	54 189	14 057 258	4 474 424
et	206 430	90 627	11 706	389	567 952	258 652
eu	349 176	157 670	4 039	114	562 739	172 598
fa	712 216	383 058	23 138	1 102	1 390 529	839 837
fi	479 830	340 425	65 714	1 464	2 514 637	1 198 436
fr	2 185 885	1 285 346	310 396	11 693	12 152 279	6 455 287
gl	161 860	72 898	12 266	484	534 529	297 054
he	261 209	126 057	24 695	359	895 104	777 238
hi	140 327	55 130	6 336	397	330 416	203 422
hr	198 670	100 743	8 451	309	462 094	246 146
hu	465 509	174 521	41 563	1 284	1 431 174	816 754
hy	264 676	181 908	19 094	902	970 970	526 903
id	524 100	225 833	33 098	1 419	1 480 566	842 242
it	1 586 855	698 980	143 021	5 406	5 894 862	3 271 852
ja	1 192 596	691 640	202 910	9 800	8 478 753	4 472 760
ka	135 333	46 280	4 919	278	261 922	159 822
kk	230 376	144 397	1 011	48	274 403	52 483
ko	486 067	170 314	24 219	978	1 125 414	725 056
la	132 258	45 476	1 563	27	128 992	66 105
lt	196 606	68 042	3 219	48	212 533	143 510
ms	335 222	76 693	10 270	457	478 597	310 896
nl	1 999 092	956 912	27 762	618	2 082 062	1 198 094
nn	151 857	44 143	4 529	121	217 312	125 392
no	529 426	253 116	23 900	946	1 239 642	691 392
pl	1 387 164	799 743	159 263	4 117	5 987 840	2 461 500
pt	1 022 524	727 228	103 004	5 462	4 932 142	2 708 120
ro	404 748	232 128	32 362	1 270	1 470 365	625 110
ru	1 602 761	972 175	212 012	7 243	8 488 280	4 564 745
sh	451 298	338 112	15 220	365	1 311 625	213 360
simple	155 887	81 648	8 775	306	429 659	274 269
sk	232 551	89 343	8 025	226	410 670	224 801
sl	167 119	64 200	7 702	331	365 134	197 270
sr	630 870	492 816	17 744	703	2 732 083	481 643
sv	3 740 411	3 123 670	135 176	33 492	20 051 166	4 207 522
ta	132 424	90 986	8 931	278	489 078	255 385
th	135 627	69 925	12 603	633	579 079	362 627
tr	343 216	163 116	22 326	1 063	1 110 192	689 826
uk	994 030	578 669	81 474	1 616	3 867 729	1 414 722
ur	154 282	114 628	3 197	182	258 602	194 405
uz	133 774	25 082	585	31	55 661	23 286
vi	1 241 487	1 053 035	41 343	1 806	2 722 811	1 600 858
vo	124 189	655	9	-	1 525	1 374
zh	1 099 744	630 425	112 429	5 187	4 974 101	2 738 099
zh-min-nan	267 615	40 192	161	2	61 887	4 898

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..

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 nearly 200 million references of Wikipedia articles in 55 language versions. Comparing results with basic extraction (see subsection 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).

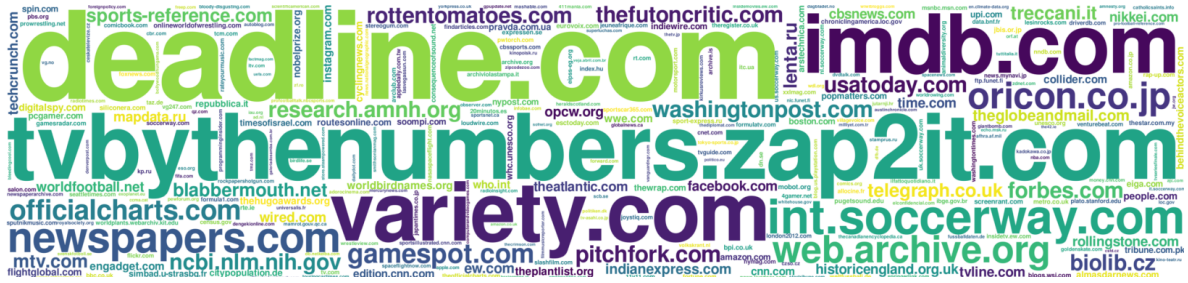

Figure 4. The most popular domains in URL of references of Wikipedia articles in 55 language versions. Source: own calculations based on Wikimedia Dumps as of March 2020 using complex extraction method.

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" [33].

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 extendent version of the results are placed on the web page: http:/data.lewoniewski.info/sources/.

Table 6 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.

Source	Language version of Wikipedia														
	ar	de	en	es	fa	fr	it	ja	nl	pl	pt	ru	sv	vi	zh
ad.nl	4 169	166	633	11 663	6 153	1 086	5 971	2 737	3	1 992	4 003	7 161	13 152	2 142	12 739
adorocinema.com	4 189	17 030	3 731	1 402	-	13 204	17 889	8 990	16 592	141	2	15 003	20 774	5 757	25 859
allocine.fr	2 051	390	929	2 138	901	2	565	1 767	2 323	1 586	1 488	963	517	4 818	4 491
almaany.com	3	23 568	5 249	27 303	391	4 592	18 098	21 354	-	-	10 374	7 209	924	13 552	32 987
appledaily.com.tw	7 260	24 734	3 917	31 354	14 794	43 411	42 064	840	-	-	4 103	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	2 174	2 255	765	2 011	3 130	1 223	1 561	2 165	1 882	1	1 109	1 818
dwdl.de	1 386	5	1 359	19 652	8 051	801	2 716	26 042	5 155	4 579	27 221	32 027	5 448	11 976	32 793
eiga.com	2 719	7 745	452	1 609	3 919	2 000	3 130	3	22 464	1 528	926	2 863	5 463	174	33
elcinema.com	1	23 353	4 628	38 243	1 744	1 585	25 524	40 045	12 266	14 817	35 232	12 767	7 341	15 563	26 656
expressen.se	1 392	557	300	1 379	8 263	389	487	6 097	505	545	973	883	2	3 011	1 724
formulatv.com	112	1 186	679	5	5 705	323	202	59 424	22 695	5 733	248	1 171	25 332	24 837	32 378
hln.be	2 052	3 577	1 817	17 379	15 411	1 471	24 548	55 133	4	2 069	5 241	17 063	24 763	4 085	4 307
ibge.gov.br	-	18 761	13 284	2 115	-	19 876	-	-	7 030	-	4	4 275	22 550	2 902	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	-	-	7 107	5	44 201	24 945	5 539	6 575
irna.ir	1 806	66 843	8 072	20 057	1	38 803	66 342	42 350	17 815	-	16 456	21 773	-	11 503	17 543
kp.ru	3 177	1 809	874	6 625	3 459	2 419	7 793	3 563	5 480	634	13 005	4	5 915	2 236	1 395
lenta.ru	352	325	462	930	1 192	480	1 254	785	2 363	166	1 342	1	1 578	310	676
lesinrocks.com	1 941	2 308	1 004	1 600	1 399	3	859	6 069	2 301	9 497	3 817	3 074	9 032	3 804	2 401
mobot.org	6 862	125 005	4 337	552	11 203	4 969	5 210	10 805	6 734	2	1 095	37 401	13 186	930	12 005
news.livedoor.com	2 529	31 803	1 628	2 967	11 697	9 632	13 447	5	-	24 057	10 329	6 965	28 944	388	98
news.mynavi.jp	1 522	5 110	1 394	12 368	4 268	15 865	16 939	4	-	40 700	3 880	11 560	7 180	410	45
nikkei.com	3 193	1 096	694	5 571	790	3 854	1 402	2	1 977	4 031	1 524	3 870	12 832	836	64
oricon.co.jp	226	360	60	167	686	121	347	1	2 606	91	131	204	1 115	22	3
regeringen.se	9 566	12 561	4 789	21 114	5 065	68 510	-	64 855	17 468	33 056	4 711	25 867	5	3 017	45 773
repubblica.it	413	205	173	260	2 403	136	3	1 188	662	348	845	407	1 221	1 064	466
research.amnh.org	49 400	49 866	16 304	13 141	-	28 287	24 255	-	14	10 293	24 065	3 317	-	2	24 727
rottentomatoes.com	16	10	5	9	18	11	19	50	44	6	9	7	109	30	14

Table 6. 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/table6

Source	Language version of Wikipedia														
	ar	de	en	es	fa	fr	it	ja	nl	pl	pt	ru	sv	vi	zh
scb.se	336	1 248	777	3 518	854	2 800	1 439	16 388	621	231	1 759	1 629	3	1 234	1 739
skijumping.pl	41 594	586	69 493	16 664	-	25 731	12 919	62 186	13 862	3	51 612	23 763	5 186	-	42 126
taz.de	3 959	3	1 648	5 397	785	692	3 821	15 993	1 918	996	13 190	1 968	2 268	577	3 684
thefutoncritic.com	139	130	19	37	87	4	16	352	335	20	40	58	458	251	82
treccani.it	333	223	278	90	2 344	59	1	2 802	229	233	75	236	1 786	871	2 809
trouw.nl	9 314	2 869	2 602	42 703	7 579	1 899	33 558	18 185	5	8 491	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	-	1 793	-	37 810	61 708	23 742	-	2 001	5
universalis.fr	5	3 273	3 525	904	6 465	8	1 223	5 180	2 512	7 534	871	2 727	1 012	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	6 504	58 271	7 212	9 972	39 942	9 639	19 417	30 018	28 707	13 486	12 178	1	9 857
volkskrant.nl	2 766	918	949	6 873	3 345	1 781	3 775	10 507	2	12 197	5 107	2 687	16 051	4 644	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 6. 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/table6

6. Classification of Sources

6.1. 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 credible.

We first analyzed values of '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 'publisher' parameter. The figure 5 shows the most commonly used publishers based on such analysis.

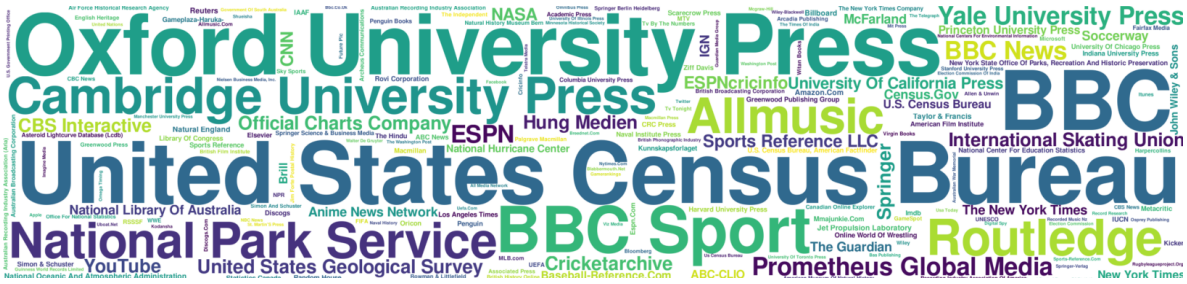

Figure 5. The most commonly used titles in 'publisher' parameter of citations templates in the references of articles in English Wikipedia in March 2020. Source: own calculations based on Wikimedia dumps using complex extraction method.

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 credibility models we assessed all journals based on the related parameter in citation templates placed in references of English Wikipedia. Table 7 shows the most popular and credible 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 place respectively. Another one of the most frequent publisher in Wikipedia - 'National Park Service' took 7th place. However it took only 94th and 58th place 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 place 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 place 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).

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

Table 7. Position in rankings of publishers in English Wikipedia depending on popularity and credibility 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/table7

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 'journal' parameter. The figure 6 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.

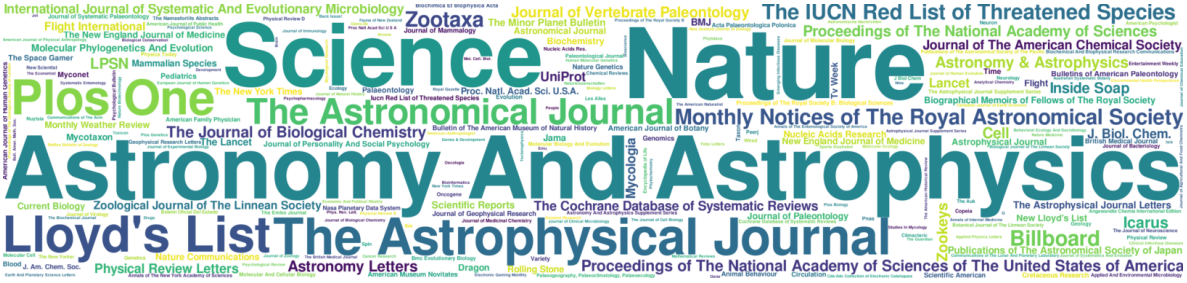

Figure 6. The most commonly used titles in 'journal' parameter of citations templates in the references of articles in English Wikipedia in March 2020. Source: own calculations based on Wikimedia dumps using complex extraction method.

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 8 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.

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

Table 8. Position in rankings of journals in English Wikipedia depending on popularity and credibility 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/table8

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 place 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 taken 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.

6.2. Semantic databases

Based on information about URL it is possible to identify title and other information related to the source. Using Wikidata and DBpedia we founded 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 9 shows position in the global ranking of the most popular and credible source with identified title based on founded items in 55 considered language versions of Wikipedia in February 2020 using different models with identified title of the source

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

Table 9. Position in the global ranking of the most popular and credible 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/table9

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 9. 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 place 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 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.

Often each of the selected items about the sources in semantic databases has additional information such as topic or type of the subject. The figure 7 shows the most popular and credibly types of the sources in 55 considered language versions of Wikipedia.

Figure 7. The most popular and credibly types of the sources in 55 considered language versions of Wikipedia based on **PR** model.

Table 10 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 has 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.

Peer-reviewed version available at Information 2020, 11; doi:10.3390/info11050263

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

Table 10. 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/table10

Based on the knowledge about type of each source we decide 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 figure 8 shows the most popular periodical sources in Wikipedia articles from 55 language versions using **PR** model. 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.

Figure 8. The most popular periodical sources in Wikipedia articles from 55 language versions in February 2020 using **PR** model. Source: own calculations based on Wikimedia Dumps using complex extraction of references.

The most popular periodical sources in Wikipedia articles from 55 language versions using different popularity and reliability models in February 2020 showed in table 11. We can observe, that there are sources, that has "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 depend on 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 of the in frequency ranking, but at the same time took 4th place in **PmR** model.

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

Table 11. 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/table11

Given local rankings of periodical we can consider the difference of credibility and popularity between different language versions. Table 12 show 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 place 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 on the 2nd place in Arabic Wikipedia, but in English, Persian, Italian, Japanese, Russian Wikipedia position if 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. To the most reliable sources in French Wikipedia includes: ”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 a 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.

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 12. 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/table12

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 12. 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/table12

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

Table 12. 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/table12

7. Different Time Periods

Using complex extraction of the references apart 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 month.

Table 13 shows position in rankings of popular and reliability sources with identified title depending on time 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.

Sources	Months			
	2019-12	2020-01	2020-02	2020-03
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

Table 13. Position in rankings of popular and reliable sources depending on time period in all considered languages 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

Next we decided to limit the list of the sources to periodical ones (as it was done in subsection 6.2). Table 14 shows position in rankings of popular and credibly sources depending on time period in all considered languages versions using **PR** model. Similarly to 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".

Sources	Months			
	2019-12	2020-01	2020-02	2020-03
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

Table 14. Position in rankings of popular and credibly sources depending on time period in all considered languages 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

Results showed, that in 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 credible 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 time period are mainly due to a large margin in absolute values of popularity and credibility 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. Next section will describe method and results of measuring

8. Growth Leaders

The Wikipedia articles may have a long edition history. Information and sources in such articles can be changed many times. More over, 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 credibility 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 credibility were changed (increase or decrease) in 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: well known and often used website in Wikipedia references dramatically lost confidence (reputation) as a credible 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 period of time.

F model showed how many references in Wikipedia articles contains specific source. 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.

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
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 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

Table 16 shows which of the sources had the largest growth of credibility 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 source 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.

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

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

9. Conclusions and Future Work

In this paper we used basic and complex extraction methods to analyze nearly 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. Using DBpedia and Wikidata we automatically identified the alignment of the sources to specific field to find the best ones in specific area. Additionally, we analyzed the differences of popularity and reliability assessment of the sources between different periods of time. Moreover, we also conducted analysis of the growth leaders in each considered months.

Sources analysis was conducted in various ways. One of the approach was to extract information from citation templates. Based on the related parameter in references of English Wikipedia we found the most popular publishers: United States Census Bureau, Oxford University Press, BBC, BBC Sport, Cambridge University Press, Routledge and others. 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 most reliable types of sources in all considered 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 provide more reliable information for Japanese and Chinese languages. In Dutch and Polish Wikipedia, archive websites have 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 is on 7th place in German Wikipedia and on 14th place in Portuguese Wikipedia, while other languages has 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.

Using information about the sources' 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 apart from 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", "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 credibility measurement.

We also provided analysis on the growth leaders in different months. Results from **PR** model showed that in December 2019 "Variety" and "Deutsche Jagd-Zeitung" were leading growing reliable sources in German Wikipedia, "Variety" and "Entertainment Weekly" – in French Wikipedia, "Lenta.ru" and "Entertainment Weekly" – 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" – in French Wikipedia, "Variety" and "Lenta.ru" – 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" – in French Wikipedia, "New York Post" and "Novosti Kosmonavtiki" – in Russian Wikipedia.

Some of extended results on reliability assessment of the sources in Wikipedia are placed in BestRef project [35].

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

References

1. Wikipedia Meta-Wiki. List of Wikipedias. https://meta.wikimedia.org/wiki/List_of_Wikipedias.
2. English Wikipedia. Reliable sources. https://en.wikipedia.org/wiki/Wikipedia:Reliable_sources.
3. Internet Live Stats. Total number of Websites. https://www.internetlivestats.com/total-number-of-websites/.
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.
5. Price, R.; Shanks, G. A semiotic information quality framework: development and comparative analysis. In *Enacting Research Methods in Information Systems*; Springer, 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 quarterly* **2013**, pp. 777–794.
7. Stvilia, B.; Twidale, M.B.; Smith, L.C.; Gasser, L. Assessing information quality of a community-based encyclopedia. *Proc. ICIQ* **2005**, pp. 442–454.
8. Blumenstock, J.E. Size matters: word count as a measure of quality on Wikipedia. Proceedings of the 17th international conference on World Wide Web. ACM, 2008, pp. 1095–1096.
9. Lucassen, T.; Schraagen, J.M. Trust in wikipedia: how users trust information from an unknown source. Proceedings of the 4th workshop on Information credibility. ACM, 2010, pp. 19–26.
10. Yaari, E.; Baruchson-Arbib, S.; Bar-Ilan, J. Information quality assessment of community generated content: A user study of Wikipedia. *Journal of Information Science* **2011**, *37*, 487–498.
11. Conti, R.; Marzini, E.; Spognardi, A.; Matteucci, I.; Mori, P.; Petrocchi, M. Maturity assessment of Wikipedia medical articles. Computer-Based Medical Systems (CBMS), 2014 IEEE 27th International Symposium on. IEEE, 2014, pp. 281–286.
12. Lewoniewski, W.; Węcel, K.; Abramowicz, W. Analysis of references across Wikipedia languages. International Conference on Information and Software Technologies. Springer, 2017, pp. 561–573.
13. Piccardi, T.; Redi, M.; Colavizza, G.; West, R. Quantifying Engagement with Citations on Wikipedia. *arXiv preprint arXiv:2001.08614* **2020**.
14. Nielsen, F.Å.; Mietchen, D.; Willighagen, E. Scholia, scientometrics and wikidata. European Semantic Web Conference. Springer, 2017, pp. 237–259.
15. Teplitskiy, M.; Lu, G.; Duede, E. Amplifying the impact of open access: Wikipedia and the diffusion of science. *Journal of the Association for Information Science and Technology* **2017**, *68*, 2116–2127.
16. Jemielniak, D.; Masukume, G.; Wilamowski, M. The most influential medical journals according to Wikipedia: quantitative analysis. *Journal of medical Internet research* **2019**, *21*, e11429.
17. Fetahu, B.; Markert, K.; Nejdl, W.; Anand, A. Finding news citations for wikipedia. Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, 2016, pp. 337–346.
18. Ferschke, O.; Gurevych, I.; Rittberger, M. FlawFinder: A Modular System for Predicting Quality Flaws in Wikipedia. CLEF (Online Working Notes/Labs/Workshop), 2012, pp. 1–10.
19. Flekova, L.; Ferschke, O.; Gurevych, I. What makes a good biography?: multidimensional quality analysis based on wikipedia article feedback data. Proceedings of the 23rd international conference on World wide web. ACM, 2014, pp. 855–866.
20. Shen, A.; Qi, J.; Baldwin, T. A Hybrid Model for Quality Assessment of Wikipedia Articles. Proceedings of the Australasian Language Technology Association Workshop 2017, 2017, pp. 43–52.

21. di Sciascio, C.; Strohmaier, D.; Errecalde, M.; Veas, E. WikiLyzer: interactive information quality assessment in Wikipedia. Proceedings of the 22nd International Conference on Intelligent User Interfaces. ACM, 2017, pp. 377–388.
22. Dang, Q.V.; Ignat, C.L. Measuring Quality of Collaboratively Edited Documents: The Case of Wikipedia. Collaboration and Internet Computing (CIC), 2016 IEEE 2nd International Conference on. IEEE, 2016, pp. 266–275.
23. Lewoniewski, W.; Węcel, K.; Abramowicz, W. Relative Quality and Popularity Evaluation of Multilingual Wikipedia Articles. *Informatics* **2017**, *4*. doi:10.3390/informatics4040043.
24. Lewoniewski, W.; Węcel, K.; Abramowicz, W. Multilingual Ranking of Wikipedia Articles with Quality and Popularity Assessment in Different Topics. *Computers* **2019**, *8*. doi:10.3390/computers8030060.
25. Warncke-wang, M.; Cosley, D.; Riedl, J. Tell Me More: An Actionable Quality Model for Wikipedia. WikiSym 2013, 2013, pp. 1–10. doi:10.1145/2491055.2491063.
26. Lih, A. Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource. *5th International Symposium on Online Journalism* **2004**, p. 31.
27. Liu, J.; Ram, S. Using big data and network analysis to understand Wikipedia article quality. *Data & Knowledge Engineering* **2018**.
28. Wilkinson, D.M.; Huberman, B.a. Cooperation and quality in wikipedia. *Proceedings of the 2007 international symposium on Wikis WikiSym 07* **2007**, pp. 157–164. doi:10.1145/1296951.1296968.
29. Kane, G.C. A multimethod study of information quality in wiki collaboration. *ACM Transactions on Management Information Systems (TMIS)* **2011**, *2*, 4.
30. Characterizing Wikipedia Citation Usage. Analyzing Reading Sessions. https://meta.wikimedia.org/wiki/Research:Characterizing_Wikipedia_Citation_Usage/Analyzing_Reading_Sessions. [Online; accessed 29-Feb-2020].
31. Lewoniewski, W. The method of comparing and enriching information in multlingual wikis based on the analysis of their quality. Phd, Poznań University of Economics and Business, 2018.
32. Lerner, J.; Lomi, A. Knowledge categorization affects popularity and quality of Wikipedia articles. *PloS one* **2018**, *13*, e0190674.
33. Wikimedia Downloads. English Wikipedia latest database backup dumps. https://dumps.wikimedia.org/enwiki/latest/.
34. English Wikipedia. 2019–20 coronavirus pandemic. https://en.wikipedia.org/wiki/2019%E2%80%9320_coronavirus_pandemic. [Online; accessed 30-Mar-2020].
35. BestRef. Popular and reliable sources of Wikipedia. https://bestref.net.