



Observatorio de la diversidad lingüística y cultural en la Internet
Observatoire de la diversité linguistique et culturelle dans l'Internet
Observatory of the Linguistic and Cultural Diversity in the Internet

<https://OBDILCI.ORG/>

WEB Multilingualism report #6 Wikimedia Multilingualism Daniel Pimienta, OBDILCI, September 2025

This series of reports has been made possible thanks to the courtesy access granted by DataProvider.com to OBDILCI. DataProvider.com maintains a database gathering information related to a series of websites close to the total WWW. This report is the only one from the series which does not need DataProvider.com inputs as it is directly based on the excellent Wikimedia's language stats.

PREVIOUS REPORTS

- The state of Multilingualism of the WWW, Technical Report #5: [Web Multilingualism analyzed by ccTLD, languages, gTLDs and much more: winners and losers](#)
- Informes sobre el multilingüismo de la Web #4: [Una evaluación de la reciprocidad en el uso mutuo del español y del portugués en las Web lusófona e hispanófona](#). 6/2025
- L'état de multilinguisme de la Toile, rapport technique #3 : [Une caractérisation du Web francophone à partir d'une série de paramètres, en comparaison avec d'autres langues dominantes sur la Toile](#). 5/2025
- The state of Multilingualism of the WWW, Technical Report #2: [Exploring web presence and multilingualism of European minority languages with associated gTLDs](#). 3/2025
- The state of Multilingualism of the WWW, Technical Report #1: [Approximation of the Rate of multilingualism of the WWW](#), 2/2025

MAIN FINDINGS

CONTENT

PREVIOUS REPORTS	1
MAIN FINDINGS	1
FOCUS OF THIS REPORT	3
METHOD, SOURCES AND BIASES	3
WIKIPEDIA	5
NUMBER OF ARTICLES PER LANGUAGE	6
ACTIVE EDITORS	8
DEPTH	9
WIKIBOOKS – NUMBER OF MODULES	10
WIKIQUOTES	11

WIKISOURCE.....	11
WIKIVERSITY.....	12
WIKTIONARY	12
WIKINEWS.....	13
WIKIVOYAGE	14
WIKIMEDIA AS A COMPREHENSIVE PACKAGE	15
A COMPARISON OF RESULTS BETWEEN ENGLISH, FRENCH, GERMAN, SPANISH AND PORTUGUESE	16
OTHER OPEN ONLINE ENCYCLOPEDIAS	17
“MINOR” LANGUAGES, MAJOR PERFORMANCES.....	21
CONCLUSION.....	23
REFERENCES.....	23

Table 1: List of Wikimedia elements.....	4
Table 2: Sorted top languages by Factor formula.....	5
Table 3: Sorted top languages by number of Wikipedia articles.....	6
Table 4: Sorted top languages by Wikipedia edits.....	7
Table 5: Sorted top languages by number of active users.....	8
Table 6: Sorted top languages by Wikipedia depth.....	9
Table 7: Sorted top languages by WikiBooks number of modules.....	10
Table 8: Sorted top languages by WikiQuote's content pages.....	11
Table 9: Sorted top languages by WikiSource's pages.....	11
Table 10: Sorted top languages by WikiVersity's modules.....	12
Table 11: Sorted top languages by Wiktionary's entries.....	12
Table 12: Sorted top languages by WikiNews entries.....	13
Table 13: Sorted top languages by WikiVoyage's articles.....	14
Table 14: Top languages by Wikimedia presence.....	15
Table 15: Sorted top languages in Wikimedia as a whole.....	15
Table 16: Comparison of results between English, French, German, Spanish and portuguese.....	16
Table 17: Online encyclopedias	18
Table 18: Major online encyclopedias	20
Table 19: Indigenous language in top active users over user indicator.....	21
Table 20: Indigenous languages in Wiktionary	22
Table 21: Number of languages per Wikimedia element sorted.....	23

FOCUS OF THIS REPORT

This report focuses the multilingualism of online encyclopedias. It dedicates the main attention on the analysis of Wikimedia language statistics and complete with a panoramic view of all online encyclopedias associated to one or more languages. The report highlights the languages which are particularly successful in Wikimedia's realm and, elaborate a comparison for main languages based on previous analysis made in 2021 in order to perceive the trends. The report shows a table of all online encyclopedias, established in 2021, and update this table with today's figures.

METHOD, SOURCES AND BIASES

The method stands on Wikimedia's language statistics. Wikimedia is probably, and from far, the most multilingual application existing in the Web, with 357 languages and, additionally the most comprehensive and trustable source of language statistics for all its components.

WIKIMEDIA COMPONENTS

Wikipedia : the most famous encyclopedia. What would be the best indicator among the set of parameters? Logically, it is the number of articles, with English, as a winner, with a little more than 7 million articles representing 10,8% of the 65 million total number of articles. The second language, in terms of articles, not so far from English, is Cebuano, a language of Philippines, with close to 21M speakers L1+L2.

A closer look to Wikipedia dynamics shows that the number of articles could be a biased indicator, mainly for two reasons, especially when they are summed up: 1) some linguistic Wikipedia's articles are made from automatic translation from another language and 2) some linguistic Wikipedia's show a very low level of updates. It is not fair to compare original articles with translated ones and still less if afterwards they stay still and unchanged for years...

The excellent stats provided in https://meta.wikimedia.org/wiki/List_of_Wikipedias offer more interesting indicators, beyond the number of articles : number of editors, number of active editors, number of edits, number of pages, number of files, number of administrators (editors named to play a special role), and, the most interesting for whoever looks for the most objective indicator, *depth*. Depth has been defined as an "indicator of the encyclopedia's collaborative quality", by the formula:

$$\text{Depth} = \text{edits} / (\text{articles} + \text{non-articles}) \times (\text{non-article/article})^2$$
 where non-articles are user pages, redirects, images, "project" pages, categories, templates, and all talk pages.

In one sense, *depth* focuses the core of Wikipedia, the very article and measure the prorated of edits per pure article as a fair reflect of the vitality of Wikipedia. However, it could favor versions with very low number of articles (say less than 20 000) and many edits; those uncommon cases provoke some noise in the indicator with gigantic figures (the top is Cree with only 13 articles and 40 000 edits leading to a depth 400 times higher than English's). If those cases are ignored, English leads, followed by Serbo-Croatian, Vietnamese, Thai, Urdu, Arabic, Bangla, French and Hebrew.

In an early version of our model for measuring the web contents per language, we used Wikimedia as an input. We created a special indicator we called *Factor* which used depth and other elements to try to obtain a fair measure of the strength of languages in Wikipedia, avoiding that languages with automatic translation and low edits could get favored and also tried to make it independent of the number of speakers (however limiting its use for languages with more than 1M speakers).

$$\text{Factor} = \text{Articles} \times \text{Edits} \times \text{Editors} \times \text{Active Editors} \times \text{Depth} / (\text{L1} + \text{L2 speakers})^2$$

In the following, factor is not computed for languages with a number of speakers less than 200 000 or number of articles less than 50 000.

In 2021, the top languages using this formula were: English, Hebrew, Italian, German, French, Swedish, Finnish, Norwegian, Dutch, Japanese, Polish, Catalan, Hungarian and Serbo-Croatian, Czech, Ukrainian, Danish, Russian, Armenian, Persian, Spanish and Portuguese.

In 2025, the list evolved into : English, Hebrew, German, Italian, Serbian, French, Norwegian, Dutch, Welsh, Vietnamese, Polish, Swedish, Japanese, Finnish, Czech, Ukrainian, Esperanto, Estonian, Basque, Serbo-Croatian and Catalan.

We will check in the analysis part of the report the classification of each relevant indicators related to Wikipedia; but there is more in Wikimedia than Wikipedia.

Here after the table showing all elements of Wikimedia and where to get the stats per language.

Table 1: List of Wikimedia elements

WIKIMEDIA ELEMENT	THEME	LANGUAGE STATS PAGE
Wikipedia	Open access encyclopedia	https://meta.wikimedia.org/wiki/List_of_Wikipedias
WikiBooks	Open access educational books	https://meta.wikimedia.org/wiki/Wikibooks/Table
WikiSource	Open access library	https://meta.wikimedia.org/wiki/Wikisource/Table
WikiQuote	Open access quotes	https://meta.wikimedia.org/wiki/Wikiquote/Table
WikiVersity	Open access educational resources	https://meta.wikimedia.org/wiki/Wikiversitye/Table
Wiktionary	Open access dictionary	https://meta.wikimedia.org/wiki/Wiktionary/Table
WikiNews	Open access news repository	https://meta.wikimedia.org/wiki/Wikinews/Table
WikiVoyage	Open access guides for travel	https://meta.wikimedia.org/wiki/Wikivoyage/Table
WikiCommons	Open access photos and other medias	
WikiData	Open access data bases	
WikiSpecies	Open access species base	
WikiFunctions	Open access functions base	
MediaWiki	Open access software base	

The study will only process the first ones which have linguistic versions and offer stats per language. The last 5 ones are somehow multilingual with main support in English.

The data proposed by Wikimedia in its linguistic stats are bias free and updated daily and this study will try to extract the substance of those stats as for the concerned languages.

We use Ethnologue for demo-linguistic data whenever appropriate.

WIKIPEDIA

The use of our factor's formula shows the following results in the first 40 top part of the ranking. Relative presence is the ratio of the factor number by the L1+L2 population. The differences of ranking are shown in the rightest column, 0 meaning no change, +n meaning a gain on ranks still 2021. Highest values are highlighted in red and lowest in blue.

Table 2: Sorted top languages by Factor formula

Language	FACTOR	%FACTOR	REL. PRES.	2025vs. 2021
English	568712	31.22%	0.46	0
Hebrew	403385	22.14%	4.07	0
German	156651	8.60%	2.27	1
Italian	81056	4.45%	2.89	-1
Serbian	78748	4.32%	7.01	
French	66442	3.65%	0.87	-1
Norwegian	54597	3.00%	12.09	-7
Dutch	37007	2.03%	8.67	1
Welsh	33512	1.84%	51.48	1
Vietnamese	29174	1.60%	1.50	-3
Polish	28766	1.58%	3.99	0
Swedish	27164	1.49%	19.74	-9
Japanese	26938	1.48%	1.19	-11
Finnish	22484	1.23%	10.71	-6
Czech	21782	1.20%	4.67	0
Ukrainian	16906	0.93%	3.57	0
Esperanto	16709	0.92%	37.41	65
Estonian	15798	0.87%	25.45	67
Basque	15231	0.84%	43.26	60
Serbo-Croatian	14846	0.81%	2.36	
Catalan	12406	0.68%	8.38	-9
Persian	12318	0.68%	1.33	-2
Danish	10669	0.59%	5.34	-6
Hungarian	7832	0.43%	4.51	-11
Russian	5915	0.32%	0.81	-7
Armenian	5830	0.32%	8.46	-7
Korean	5735	0.31%	0.89	-3
Icelandic	5563	0.31%	18.61	93
Macedonian	4039	0.22%	9.38	70
Latvian	2992	0.16%	6.80	69
Breton	2960	0.16%	43.25	79
Turkish	2904	0.16%	0.72	-8
Spanish	2528	0.14%	0.37	-12
Slovenian	2187	0.12%	8.59	57
Lithuanian	1986	0.11%	8.04	52
Portuguese	1863	0.10%	0.44	-14
Arabic	1806	0.10%	0.38	-9
Luxembourgish	1493	0.08%	12.39	79
Asturian	1457	0.08%	19.40	57
Bulgarian	1062	0.06%	3.87	15
Chechen	992	0.05%	34.16	new
Galician	960	0.05%	6.69	17

What does this table tell us?

- English and Hebrew are solid first places if one looks at the percentages (the percentage is made by dividing the factor by the total sum of factors), summing up to 50% between both.

- European languages trust the first places with the exception of Vietnamese, Japanese, Esperanto, Persian, Korean, Turkish, Arabic.
- Not many changes in the top of ranking except Norwegian, Swedish, Japanese and Finnish have lost several ranks.
- Some languages have boosted their rank since 2021: Esperanto, Estonian, Basque, Icelandic, Macedonian, Latvian, Breton; Slovenian, Lithuanian and Luxembourgish.
- Strongest drop of ranking for Portuguese and Spanish.
- Local languages of Europe in the top ranks: Welsh, Basque, Catalan, Breton and Asturian.
- The *relative presence* indicates the number of Wikipedia articles per speaker, with Welsh, Breton and Esperanto as higher figures and Spanish, Arabic, Portuguese and English as the lowest.

Note that Serbia and Serbo-Croatian comparison is not available as in 2021 we joined the figures of all Serbo-Croatian macro-language.

Let's see now the other rankings, keeping in mind their definitions and limitations.

NUMBER OF ARTICLES PER LANGUAGE

Table 3: Sorted top languages by number of Wikipedia articles

TOTAL	65 380 658	100%	VS. 2021
INCREASE vs. 2021			14.5%
English	7 042 461	10.77%	0
Cebuano	6 116 183	9.35%	0
German	3 043 035	4.65%	1
French	2 703 580	4.14%	2
Swedish	2 615 004	4.00%	-2
Dutch	2 194 947	3.36%	1
Russian	2 059 922	3.15%	1
Spanish	2 055 799	3.14%	2
Italian	1 931 384	2.95%	0
Polish	1 666 455	2.55%	1
Egyptian Arabic	1 628 639	2.49%	
Chinese	1 495 506	2.29%	0
Japanese	1 470 351	2.25%	1
Ukrainian	1 388 448	2.12%	2
Vietnamese	1 295 892	1.98%	1
Arabic	1 276 591	1.95%	
Waray	1 266 757	1.94%	-1
Portuguese	1 152 989	1.76%	0
Persian	1 050 644	1.61%	1
Catalan	779 651	1.19%	2
Indonesian	739 512	1.13%	-2
Korean	719 155	1.10%	2
Serbian	709 869	1.09%	
Norwegian	654 777	1.00%	-3

Turkish	644 720	0.99%	-5
Chechen	602 280	0.92%	-6
Finnish	600 670	0.92%	-3

- The growth of the total number of articles since 2021 is 14.5%.
- Not many changes in the ranking (Arabic and Serbian are not compared with 2021 because in the 2021 run we worked with macro-languages and regroup all data for macro-languages components.
- Some good results in number of articles are not reflected in the previous "factor" table because their "process performance" are too weak (Cebuano, Arabic, Waray, Indonesian) as for the factor equation.

Number of article's edition per language. We add another indicator probably more significative for comparisons, the average number of edit per article.

Table 4: Sorted top languages by Wikipedia edits

Language	Edits	Growth	LANGUAGE	Edit/article
English	1 302 182 355	0	English	184,9
German	257 156 802	0	Hebrew	109,2
French	227 811 660	0	Serbo-Croatian	92,1
Spanish	168 619 075	0	German	84,5
Russian	147 675 475	1	French	84,3
Italian	146 037 380	-1	Spanish	82,0
Japanese	105 894 111	0	Italian	75,6
Chinese	88 507 247	0	Japanese	72,0
Polish	77 229 592	1	Russian	71,7
Vietnamese	73 731 126	-1	Thai	71,0
Arabic	71 674 448	1	Portuguese	61,2
Portuguese	70 522 312	-1	Chinese	59,2
Dutch	69 606 069	0	Vietnamese	56,9
Swedish	57 823 399	-1	Korean	56,2
Ukrainian	45 910 269	1	Arabic	56,1
Serbo-Croatian	42 470 875	-2	Turkish	55,5
Persian	42 372 774	0	Hungarian	50,5
Hebrew	41 611 774	1	Welsh	49,8
Korean	40 421 763	1	Malayalam	49,4
Turkish	35 808 840	2	Bangla	48,2
Catalan	35 783 856	0	Polish	46,3
Cebuano	35 781 802	-4	Catalan	45,9
Serbian	30 328 108		Greek	43,7
Hungarian	28 285 736		Czech	43,7
Indonesian	27 672 844		Serbian	42,7
Norwegian	25 260 574		Bulgarian	41,4
Czech	25 117 375		Persian	40,3
Finnish	23 448 107		Luxembourgish	40,1

English leads in both tables and German, French, Japanese, Spanish, Italian and Russian keep top positions in both tables. Hebrew and Serbo-Croatian gain many ranks in the second table,

as well as Thai. In comparison with 2021, there is a global growth of 25%, close to the double of the growth of articles, which is a sign of vitality, and not much changes in the rankings. The fact to have filtered out the results of languages with low number of articles or low number of speakers prevents to see at first glance some singularities, such as Cree has only 13 articles but close to 41 000 edits and Arakanese have 0 registered articles but more than 1000 edits and more than 2000 active editors (labeled as “active users” in Wikimedia terminology).

ACTIVE EDITORS

Active editors have grown globally 27% since 2021. We have added, for a more meaningful comparisons, the ratio of active users over L1+L2 speakers, multiplied by 1000 and the ratio users over active users (filtered by number of articles higher than 50 000).

Table 5: Sorted top languages by number of active users

Language	Act. Users	Gr.	Language	1000*Act. /Pop.	LANGUAGE	Act./Usr
English	107458	0	Hebrew	0,89	Occitan	13,03
German	38598	0	Icelandic	0,50	Japanese	12,02
French	38502	0	Estonian	0,43	Tamil	10,13
Japanese	28667	1	Breton	0,40	Korean	8,40
Chinese	17522	3	Norwegian	0,39	German	8,25
Spanish	11768	-3	Czech	0,37	Ukrainian	8,01
Persian	11164	3	Dutch	0,32	Persian	7,85
Polish	10848	4	Esperanto	0,31	Polish	7,70
Hebrew	8337	7	Danish	0,30	French	7,27
Dutch	8164	3	German	0,29	Hebrew	6,80
Portuguese	7782	-4	Basque	0,26	Swahili	6,72
Korean	7729	6	Polish	0,26	Czech	6,40
Russian	7687	-7	Finnish	0,24	Dutch	5,70
Arabic	7628	-3	Japanese	0,23	Hausa	5,44
Indonesian	6843	-1	Welsh	0,23	Serbian	4,63
Italian	6642	-7	Serbian	0,21	Chinese	4,61
Ukrainian	6586	0	Luxembourgish	0,18	Indonesian	4,35
Turkish	5459	-4	Asturian	0,17	Ladin	3,59
Czech	4577	1	Ukrainian	0,17	Cantonese	3,39
Vietnamese	3047		Latvian	0,14	Danish	3,27
Arakanese (*)	2278		Persian	0,14	Turkish	3,19
Serbian	2155		Swedish	0,13	Norwegian	3,17
Norwegian	2105		Slovenian	0,13	Vietnamese	3,04
Swedish	1751		French	0,12	Azerbaijani	2,86
Danish	1718		Macedonian	0,11	Arabic	2,76
Hindi	1666		Lithuanian	0,11	Armenian	2,59
Finnish	1358		Armenian	0,11	Thai	2,50
Simple English	1336		Hungarian	0,10	Italian	2,49
Hungarian	1273		Western Frisian	0,10	Portuguese	2,38
Thai	1268		Catalan	0,10	Finnish	2,23
Cantonese	1027		Italian	0,10	Minangkabau	2,18
Bangla	1008		Korean	0,10	Latvian	2,18
Catalan	938		Silesian	0,09	English	2,17
Azerbaijani	902		Azerbaijani	0,09	Hungarian	2,14
Greek	836		Bulgarian	0,07	Estonian	2,04
					Russian	2,03
					Bangla	2,02

Malay	626			Galician	0,07	Hindi	1,93
Bulgarian	589			English	0,07	Central Kurdish	1,85
Swahili	541			Occitan	0,07	Kyrgyz	1,84
Slovak	437			Irish	0,07	Swedish	1,82
Estonian	425			Belarusian	0,07	Greek	1,82
Croatian	424			Croatian	0,07	Punjabi	1,80
						Catalan	1,80

It is interesting to see that the first place of English is mainly related to its large number of speakers, as in the second table it appears in the lower part and in the middle in the third table. As for Chinese, its ratio is very low (0.01) in the bottom of the second but in the top part of the third.

DEPTH

As for the depth, 2025 shows a slight decline if the total of values of 2%.

Table 6: Sorted top languages by Wikipedia depth

Language	Depth
English	1325
Serbo-Croatian	749
Vietnamese	746
Thai	334
Urdu	310
Arabic	286
Bangla	282
French	272
Hebrew	267
Hindi	249
Malayalam	218
Chinese	214
Portuguese	205
Spanish	194
Italian	193
Turkish	187
Serbian	174
Korean	168
Russian	165
Sinhala	158
Persian	154
Tagalog	144
Indonesian	135
Romanian	126
Kannada	108
German	93
Japanese	92

This table (filtered y low number of articles or low number of speakers) shows the most collaborative languages in Wikipedia.

WIKIBOOKS – NUMBER OF MODULES

We are not able to show comparisons with 2021 as it seems that in 2021 we kept the number of edits instead of the number of modules. Strangely enough, the number of languages present in WikiBooks is now 78, lower of 38 compared with 2021. It seems that WikiBooks languages with almost no content has been moved into incubator or being archived.

Table 7: Sorted top languages by WikiBooks number of modules

LANGUAGES	MODULES	PERCENT
English	98064	33,1%
Vietnamese	51385	17,4%
Hungarian	44491	15,0%
German	33582	11,3%
French	21339	7,2%
Italian	19196	6,5%
Japanese	16887	5,7%
Portuguese	13953	4,7%
Spanish	9674	3,3%
Dutch	9520	3,2%
Polish	6925	2,3%
Indonesian	5454	1,8%
Finnish	5030	1,7%
Hebrew	4785	1,6%
Chinese	3791	1,3%
Persian	3557	1,2%
Azerbaijani	3405	1,2%
Albanian	3282	1,1%
Russian	3245	1,1%
Lithuanian	3237	1,1%
Catalan	2265	0,8%
Bangla	2214	0,7%
Basque	2187	0,7%
Thai	1600	0,5%
Czech	1564	0,5%
Danish	1338	0,5%
Korean	1314	0,4%
Hindi	1273	0,4%
Sinhala	1256	0,4%
Bashkir	1216	0,4%
Swedish	1209	0,4%

WIKIQUOTES

There are 73 languages with WikiQuote's entry, 13 more than in 2021, and the growth of content pages is 48%. To be noted, the huge growth of Estonian, and the notable growth of Azerbaijani, Hebrew, Esperanto and Arabic.

Table 8: Sorted top languages by WikiQuote's content pages

LANGUAGES	CONTENT PAGES	PERCENT	GROWTH
English	59498	16,9%	0
Italian	53477	15,2%	0
Polish	30756	8,8%	0
Russian	17405	5,0%	0
Czech	14949	4,3%	0
Estonian	13642	3,9%	33
Portuguese	11959	3,4%	1
Ukrainian	11546	3,3%	5
Hebrew	10647	3,0%	7
French	9802	2,8%	1
Persian	9504	2,7%	-5
Spanish	8362	2,4%	-2
German	7922	2,3%	-6
Serbian	5845	1,7%	-4
Turkish	5815	1,7%	-1
Azerbaijani	5489	1,6%	9
Esperanto	5236	1,5%	7
Slovak	4523	1,3%	-5
Catalan	4158	1,2%	-1
Bosnian	4157	1,2%	new
Chinese	3972	1,1%	-1
Finnish	3416	1,0%	-5
Arabic	3343	1,0%	7

WIKISOURCE

There are 90 languages in 2025 versus 70 in 2021. The total growth of pages is 109%. Many notable differences in the rankings, with growth of Hebrew, Ukrainian and Spanish and new languages in the top (Bangla and Hindi).

Table 9: Sorted top languages by WikiSource's pages

LANGUAGES	ALL PAGES	PERCENT	GROWTH
English	4701364	21,79%	1
French	4522270	20,96%	-1
Hebrew	1660283	7,69%	7
Polish	1267380	5,87%	1
Chinese	1157138	5,36%	-2
Russian	1120843	5,19%	0

Italian	835547	3,87%	1
German	651563	3,02%	-1
Bangla	635952	2,95%	new
Tamil	613238	2,84%	-1
Ukrainian	498541	2,31%	12
Spanish	318212	1,47%	23
Arabic	248653	1,15%	-1
Multilingual	236697	1,10%	-1
Swedish	203049	0,94%	-4
Telugu	196298	0,91%	2
Hindi	181573	0,84%	new
Sanskrit	162334	0,75%	-2
Portuguese	161138	0,75%	-2
Armenian	146341	0,68%	-4
Norwegian	111186	0,52%	-6
Belarusian	108653	0,50%	-2

WIKIVERSITY

WikiVersity had 13 language's versions in 2021 and grows to 18, while the number of learning modules grows 27%. German keeps leading over English and French.

Table 10: Sorted top languages by WikiVersity's modules

LANGUAGES	MODULES	PERCENT	CHANGES
German	75441	44,9%	0
English	39375	23,5%	0
French	17037	10,1%	0
Chinese	7800	4,6%	new
Italian	5359	3,2%	2
Czech	4529	2,7%	-1
Portuguese	4420	2,6%	-1
Russian	4228	2,5%	new
Multilingual	3421	2,0%	new
Spanish	2050	1,2%	-2
Slovenian	891	0,5%	-1
Arabic	751	0,4%	new
Swedish	650	0,4%	-2
Finnish	629	0,4%	-6
Greek	446	0,3%	new
Hindi	335	0,2%	new
Korean	313	0,2%	new
Japanese	208	0,1%	-5

WIKTIONARY

The 2021 edition held 141 language's version; the figures climbs to 174 with a total growth of entries of 46%. Hereafter the top languages.

Table 11: Sorted top languages by Wiktionary's entries

LANGUAGES	ENTRIES	PERCENT	CHANGES
English	8528207	18,9%	0
French	6672718	14,8%	1
Malagasy	5370032	11,9%	-1

Chinese	2154371	4,8%	0
Thai	1715993	3,8%	25
Greek	1562253	3,5%	8
Russian	1434141	3,2%	-1
German	1199696	2,7%	0
Kurdish	1004768	2,2%	3
Dutch	981360	2,2%	-1
Swedish	973610	2,2%	-1
Spanish	942160	2,1%	-5
Serbo-Croatian	924609	2,0%	-7
Polish	818257	1,8%	-3
Turkish	676323	1,5%	6
Finnish	631668	1,4%	1
Lithuanian	618981	1,4%	-4
Catalan	607351	1,3%	0
Italian	581381	1,3%	-4
Hungarian	554983	1,2%	-1
Portuguese	500036	1,1%	76
Japanese	448711	1,0%	5

Huge growth of Portuguese and Thai, important growth of Greek and some decline of Serbo-Croatian and Spanish. Malagasy loses second place to French, Kurdish gain 3 ranks and shows the interest beyond the usual main languages of Wikimedia. Quite many changes due to what looks like a very dynamic part of Wikimedia.

WIKINews

This is the segment of Wikimedia which presents the most drastic changes since 2021; it shows a growth higher than 11 times in terms of news, and this is explained by a singularity related to Russian which holds now close to 85% of the news, compared to the 12% back in 2021. Just behind Russian appears Serbian with an important growth.

Table 12: Sorted top languages by WikiNews entries

LANGUAGES	NEWS	PERCENT	CHANGE
Russian	1496851	84,8%	2
Serbian	53109	3,0%	6
Portuguese	34813	2,0%	15
French	24139	1,4%	-2
Polish	22684	1,3%	-1
English	22166	1,3%	-5
Chinese	18338	1,0%	0
German	14218	0,8%	11
Italian	12396	0,7%	14
Spanish	12284	0,7%	-5
Arabic	9159	0,5%	-4
Czech	9011	0,5%	-3

Dutch	4836	0,3%	16
Catalan	4799	0,3%	-4
Japanese	3995	0,2%	15
Greek	3065	0,2%	-3
Tamil	3052	0,2%	new
Limburgish	2629	0,1%	new
Swedish	2077	0,1%	-7
Ukrainian	1982	0,1%	-6

The analysis of the Russian singularity led to the conclusion that the reason of this situation is 1) that the alternative media group NewsRU.com, quite critical with officiality in Russia, closed in May 2021 and all its archive were moved into WikiNews. Additionally, WikiNews was heavily fed afterwards in Russian, from probably the same journalists (there is 57 active users), the NewsRU archive representing half of the Russian news and the other half being fresh news after 2021. Something similar seems to happen with Serbian news, in relation with alternative Serbian Beta Press agency. In conclusion, WikiNews seems to have transformed into a space of resistance for Russian and Serbian medias.

Besides this extreme singularity, this is the segment with the most important changes since 2021, with a stop of English (only 2% growth) and the huge growth of news in Portuguese, German, Italian, Dutch and Japanese.

WIKIVOYAGE

WikiVoyage grows from 17 languages in 2021 to 27 in 2025, with 39% more articles and no many changes in ranking except the coming of new languages.

Table 13: Sorted top languages by WikiVoyage's articles

LANGUAGES	ARTICLES	PERCENT	CHANGES
English	33636	23,7%	0
German	20686	14,6%	0
Polish	13354	9,4%	3
Italian	12995	9,1%	0
French	9411	6,6%	0
Persian	8743	6,2%	-3
Russian	6878	4,8%	0
Chinese	5559	3,9%	2
Dutch	4332	3,0%	-1
Portuguese	4016	2,8%	-1
Spanish	3554	2,5%	1
Hebrew	2450	1,7%	0
Bangla	2040	1,4%	new
Finnish	1800	1,3%	new
Swedish	1614	1,1%	-1
Japanese	1521	1,1%	new
Indonesian	1501	1,1%	new
Vietnamese	1485	1,0%	-5

Greek	1439	1,0%	-4
Esperanto	1374	1,0%	new
Ukrainian	1095	0,8%	-4
Romanian	919	0,6%	-6
Czech	528	0,4%	new
Turkish	481	0,3%	new
Pashto	320	0,2%	new
Hindi	190	0,1%	new
Shan	185	0,1%	new

WIKIMEDIA AS A COMPREHENSIVE PACKAGE

It could be interesting to check the performance of languages in Wikimedia taken as a whole. The methodology we are using is first to identify the group of languages which are present in all Wikimedia themes, those which are present in all but one or all but 2.

From this pre-selection, we have computed the average percentage of presence in all Wikimedia segments and use this value to rank languages. Given the extreme singularity of Russian in WikiNews (85% of total news) we propose the results with and without WikiNews.

Table 14: Top languages by Wikimedia presence

PRESENCE IN ALL	PRESENCE IN ALL BUT 1	PRESENCE IN ALL BUT 2
Chinese	Arabic	<i>Bangla</i>
Czech	Dutch	<i>Catalan</i>
English	Esperanto	<i>Indonesian</i>
Finnish	Hebrew	<i>Limburgish</i>
French	Hindi	<i>Norwegian</i>
German	Korean	<i>Serbian</i>
Greek	Persian	<i>Slovenian</i>
Italian	Polish	<i>Tamil</i>
Japanese	Romanian	<i>Turkish</i>
Portuguese	Ukrainian	<i>Vietnamese</i>
Russian		
Spanish		
Swedish		

WIKIMEDIA RANKINGS

Table 15: Sorted top languages in Wikimedia as a whole

RANKING WITH WIKINEWS			RANKING WITHOUT WIKINEWS	
English	20,92%	1	English	22,71%
Russian	9,74%	2	German	9,95%
German	9,19%	3	French	7,98%
French	7,43%	4	Italian	4,85%
Italian	4,50%	5	Hebrew	3,78%

Hebrew	3,47%	6	Polish	3,40%
Polish	3,22%	7	Russian	2,92%
Chinese	2,71%	8	Chinese	2,86%
Vietnamese	2,51%	9	Vietnamese	2,74%
Spanish	2,13%	10	Spanish	2,26%
Portuguese	2,02%	11	Japanese	2,08%
Japanese	1,93%	12	Portuguese	2,03%
Dutch	1,57%	13	<i>Dutch</i>	1,69%
Persian	1,47%	14	<i>Persian</i>	1,60%
Serbian	1,13%	15	<i>Ukrainian</i>	1,20%
Czech	1,11%	16	<i>Swedish</i>	1,19%
Ukrainian	1,11%	17	<i>Czech</i>	1,17%
Swedish	1,10%	18	Arabic	0,96%
Arabic	0,92%	19	Serbian	0,96%
Finnish	0,77%	20	Finnish	0,83%
Indonesian	0,74%	21	Indonesian	0,80%
Turkish	0,71%	22	Turkish	0,78%
Bangla	0,69%	23	Bangla	0,75%
Korean	0,62%	24	Korean	0,68%
Greek	0,58%	25	Greek	0,62%
Catalan	0,57%	26	Norwegian	0,62%
Norwegian	0,57%	27	Catalan	0,60%
Tamil	0,43%	28	Tamil	0,45%
Esperanto	0,41%	29	Esperanto	0,44%
Hindi	0,37%	30	Hindi	0,41%
Romanian	0,30%	31	Romanian	0,32%
Slovenian	0,21%	32	Slovenian	0,23%
Limburgish	0,12%	33	Limburgish	0,12%

The two top languages, behind **English**, the indisputable leader of Wikimedia, are **German** and **French**; they are followed by **Italian**, **Hebrew** and **Polish**. Behind those 6 top languages, between 2 and 3% of contents, are Russian, Chinese, Vietnamese, Spanish, Japanese and Portuguese, demonstrating that internationalization efforts of Wikipedia have brought some visible results. Swedish have lost ranks in almost every segment of Wikimedia and dropped from the third group it belonged in 2021, to join Dutch, Persian, Ukrainian and Czech.

A COMPARISON OF RESULTS BETWEEN ENGLISH, FRENCH, GERMAN, SPANISH AND PORTUGUESE

Table 16: Comparison of results between English, French, German, Spanish and Portuguese

WIKIMEDIA	TOT.	SP	PO	FR	GE	EN
General indicator	343	10	12	3	2	1
Rank - %	100%	2,26%	2,03%	7,98%	9,95%	22,7%
Wikipedia -	343	8 (+2)	18 (0)	4 (+2)	3 (+1)	1 (=)
Articles	100%	3,14%	1,76%	4,14%	4,65%	10,77%
Wikipedia -	343	33 (-12)	36 (-14)	6 (-1)	3	1
Factor	100%	0,14%	0,10%	3,65%	8,60%	31,22%

Wikipedia - users	343	2	7	3	4	1
Wikipedia Active users	343	6 (-3) 3,1%	11 (-4) 2%	3 (=) 10,0%	2 (=) 10,0%	1 (=) 27,9%
Wikipedia Active/users	343	181	114	53	42	123
Wikipedia Actives/Pop.	343	194	175	93	59	124
Wikipedia Depth	343	32	29	17	43	1
Wikipedia - Edits	343	4	12	3	2	1
Wikipedia Edits/Art.	343	31	17	16	15	5
Wiktionary	174	12 (-5) 2,1%	21 (+76) 1,1%	2 (+1) 14,8%	8 (=) 2,7%	1 (=) 27,9%
WikiSource	81	12 (+23) 1,47%	19 (-2) 0,75%	2 (-1) 20,96%	9 (-1) 1,02%	1 (+1) 21,79%
WikiBooks	78	9 3,3%	8 4,7%	5 7,2%	4 11,03%	1 33,01
WikiQuote	74	12 (-2) 2,4%	7 (+1) 3,4%	10 (+1) 2,8%	13 (-6) 2,3%	1 (=) 16,9%
WikiNews	31	10 (-5) 0,7%	3 (+15) 2,0%	4 (-2) 1,4%	8 (+11) 0,8%	6 (-5) 1,3%
WikiVersity	18	10 (-2) 1,2%	7 (-1) 2,6%	3 (=) 10,1%	1 (=) 44,9%	2 (=) 23,5%
WikiVoyage	18	11 (-5) 2,5%	10 (+15) 2,8%	5 (-2) 5,6%	2 (=) 14,6%	1 (=) 23,7%

English leads in all Wikimedia elements except WikiVersity and WikiNews. If the average ranking is used instead of the average percentage French is second to English and German third. The 3 languages on the top show no notable changes since 2021; Portuguese displays more growth than decrease at the opposite of Spanish which shows more decrease than growth.

OTHER OPEN ONLINE ENCYCLOPEDIAS

In the last update of our methodology for measuring the repartition of web contents per language, we have discarded the wonderful stats of Wikimedia as input to our model, not for any doubt about their reliability, but only because they were not the reflect of the reality of web contents per language. Prior to take that decision we tried to balance Wikimedia figures with the rest of online encyclopedias, in a last attempt to obtain a fair reflection of online contents. We did an exhaustive inventory of online encyclopedias and search for their number of articles. We reached the conclusion that encyclopedias cannot be used as representation of languages online because their weight was strongly depending on languages, in no way proportionately to their number of speakers nor their online strength. It remains that we have this table extracted from (Pimienta and al., 2023) that we try to update today (in red the updates).

Table 17: Online encyclopedias¹

LANGUAGE	ENCYCLOPEDIA	NUMBER OF ARTICLES (Millions)	OTHER INFORMATION
Various	Encyclopedia of Life (EOL)	0.75 (2010) 1.9 2025 : 21,289, 852 trait and attribute records	Supported languages: Arabic, Brazilian Portuguese, English, Finnish, French, Macedonian, Piedmontese, Traditional Chinese and Turkish Interface languages: the same plus German, Spanish, Dutch, Turkish and Ukrainian. Most of the common or vernacular taxon names in EOL are from the holdings of the wikidata, which include just over a million names in 466 languages
Various	thefreedictionary.com/ Free with ads or paid	no statistics No change	English, Spanish, German, French, Italian, Chinese, Portuguese, Dutch, Norwegian, Greek, Arabic, Polish, Turkish, Russian, Hebrew It is not clear if they are parallel version or language specific.
Various	en.metapedia.org/ neonazi version of wikipedia	Marginal (5000 articles in English) 15000 16 languages NO ANSWER	Czech, Danish, German, Spanish, English, Hungarian, Dutch, Portuguese, Romanian, Slovenian, Swedish, Estonian, Croatian, Icelandic, Norwegian, Macedonian
Chinese	Baidu Baike	24.5 29.9	194 million edits 276 M 7.5 million publishers 8M
Chinese	Baik (Hudong)	18 No update found	5.8 million publishers (2013)
Chinese	SogouBaik	???? Few millions?	
Arab	Marefa	0.136636 0.161	2.4 million pages
Arab	Mawdoo3	0.15 0.30	45 (2018)
Bengali and English	bengaldia	0.0057 No update found	1450 publishers
Catalan	www.enciclopedia.cat/	0.350	
Croatian	enciklopedija.hr	0.067 No update found	Print version data
Croatian	proleksis.lzmk.hr	0.062 No update found	
Danish	Den Store Danske	0.161 0.245	1100 publishers 1 million users

¹ The main source is https://en.wikipedia.org/wiki/List_of_online_encyclopedias. As for the figures, many sources are used if they are not found directly in the site.

Dutch	winklerprins.com	0.0115 0.014	by subscription
English	britannica.com		limited free access
English Korean Chinese	Everipedia Articles copied from Wikipedia Link now to https://iq.wiki , a crypto currency encyclopedia	? Thousands	7000 active publishers (2019) 3M users (2017) free access but also blockchain market
English	Citizenship	0.017 0.024 (2023)	Statistics stopped in 2014 close to stopping Have not stopped but change owner
English	Conservapedia	0.0518 0.058	800 million pageviews 1.5 million edits https://www.conservapedia.com/Special:Statistics
English	Scholarpedia Peer-reviewed expert articles on hard science	0.0018 Very low growth	Marginal digits
English	Encyclopedia.com	0.3 0.35	Formal encyclopedia aggregator
English	Colombia Encyclopedia	0.051	Aggregated by Encyclopedia.com
English	digitaluniverse.net	No stats available	Offline Online now but not clear if frequently updated
French	Larousse	0.317 0.09	
German	retro bib	0.3 0.35	
German	https://brockhaus.de/ecs/enzy	0.3	
Hebrew English	Hamichlol	0.28 0.366	Censored version of Wikipedia for a hyper-religious audience
Korean	https://namu.wiki	6.7M	10 times more entries than Korean Wikipedia
Korean	http://rigvedawiki.net/		In 2015 most use migrate to NamuWiki
Korean	Doopedia	0.588 0.760	
Malay Sundanese Javanese	Superpedia	0.02	Site not reachable
Italian	Treccani	0.9 No stats available	Wonder if the 2021 figure is correct? Read about 420K “words”,
Lithuanian	https://www.vle.lt/		Site does not respond
Malayalam	Saravaijnanakosam	0.007	Site does not respond
Marathi	Viswakosh	0.016 0.016	
Norwegian Bokmal and Nynorsk	Store norske leksikon	0.2 (2019) No new data	3M users/month read 500K articles
Polish	https://encyklopedia.interia.pl/	0.12 (2006) No new data	

Polish	encyclopedia.pwn.pl	0.08 No new data	
Russian	Great Russian Encyclopedia	0.012 (2016) No new data	
Russian	Krugosvet	0.012 No new data	
Russian and Russia's languages	ruwiki.ru		Launched in 2023. A Kremlin compliant fork version of Russian Wikipedia
Spanish	https://www.ecured.cu/ Cuban	0.237 0.278	73,000 active 537 publishers
Spanish	Encyclonet	0.185 No new data	
Spanish	enciclopedia.us.es/	0.053	https://wikiapiary.com/wiki/ Site does not respond
Swedish	ne.se/	0.26 (2005) No new data	
Tamil	not online https://www.tamilvu.org/en/content/encyclopedia-project		Seems empty
Turkish	Eksi Sozluk	8M admissions in 2009 No new data	400,000 users 110,000 publishers 4M new admissions/year in 2013 Open for publication, each entry is kept after moderation.
Vietnamese	seems to have disappeared		Go to archive.org https://bachkhoatoanthu.vass.gov.vn

From that list the most notable online encyclopedias, in terms, of number of articles are the following:

Table 18: Major online encyclopedias

NAME	LANGUAGE	NUMBER OF ARTICLES
WIKIPEDIA	343	Total : 65M - English : 7M
BAIDU-BAIKE	Chino	~30M
BAIKE-HUDONG	Chino	>18M
EKSI-SOZLUK	Turkish	>8M
NAMUWIKI	Korean	6.7M
SOIGU-BAIKE	Chino	Few M
EOL	English + 466 ²	> 1.9M
TRECCANI	Italian	0.900M ³
DOOPEDIA	Korean	0.760M

² The 466 figure seems to come from the combination of supported machine translations and multilingual species names, not from human-written content in each language. Then Wikimedia remains the leader in multilingualism.

³ This figure sums up: the Enciclopedia Italiana articles, Specialized encyclopedias (e.g. Storia, Scienze, Arte, Diritto), Dictionaries and linguistic resources, Biographical dictionaries (Dizionario Biografico degli Italiani), the Vocabolario Treccani della Lingua Italiana, news, essays, and curated portals.

“MINOR” LANGUAGES, MAJOR PERFORMANCES

It would not be fair to conclude this study of Wikimedia’s multilingualism without a glance to the many languages we have left aside in our formula and most tables, either because of a low number of speakers or a low number of Wikipedia’s articles. Many languages have managed to get spaces in one or more of Wikimedia’s element in a manner that the statistician qualifies of “disproportionate” but that could also be qualified from the own language perspective of “bright”. Let’s dedicate a section to those languages.

In Wikipedia’s number of articles, **Cebuano**, a language of Philippines, with 6.5 million speakers is second to English with the double of articles of the third one, German. **Egyptian Arabic** is in position 11 and **Waray-waray**, another language of Philippines, with 2.6 million speakers is 17th. **Catalan** (20th), **Basque** (32nd), **Welsh** (42nd), **Galician** (53th), **Ladin** (58th) and **Asturian** (67th) are the first European minority languages. **Afrikaans** (69), **Swahili** (74) and **Malagasy** (75) are the first languages from Africa.

In terms of number of edits per article, few indigenous languages are above English :

- **Arakanese** (also known as Rakhine, a language of Myanmar, with 2M speakers), have 0 articles but 3500 edits.
- **Cree**, an indigenous language of Canada, with 42K speakers) has 41K edits for 13 articles.
- **Colognian**, (also known as Kölsch) a dialect of the City of Cologne with a quarter million speakers, has 1.3M edits for 303 articles.
- **Kalaallisut**, the official language of Greenland, of the Inuit family, spoken by 57K people, has 76K edits for 247 articles. Of the same family, **Inuktitut**, indigenous of Canada, with 42K speakers is not far with 48K edits for 428 articles.

Not far, behind English, **Tigré**, an indigenous language of Eritrea, with 900K speakers; has 6300 edits for 39 articles; **Aromanian**, an indigenous language from Albania, 209K edits for 1388 articles and **Dzongkha**, the national language of Bhutan with 315K speakers shows 3700 edits for 323 articles.

In active users per users, there are 30 of those languages prior to Occitan, and many more after, as it seems that this ratio tends to be high when the number of users is low. Let’s mention the first ones (which were filtered out for low number of articles). Note that the average value of this indicator is 6.25 and the median is 1.61.

Table 19: Indigenous language in top active users over user indicator

LANGUAGE	SPEAKERS	COUNTRY	STATUS	Act. users/usr
Nias	0.9 M	Indonesia	Stable	360.00
Northern	13.7M	South Africa	Official	282.05
Arakanese	2.0M	Myanmar	Wide	94.34
Pa'O	0.9M	Myanmar	Stable	81.97
Crimean Tatar	0.6M	Ukraine	Stable	68.23
Saraiki	26M	Pakistan	Stable	61.45
South Ndebele	2.5M	South Africa	Official	39.72
Gorontalo	0.5M	Indonesia	Endangered	35.67
Tumbuka	2.3M	Malawi	Stable	31.87
Mossi (Moore)	11.9M	Burkina	Wide	28.41
Nupe	2.0M	Nigeria	Wide	23.64

Observing the indicator active users per population, 20 languages appears on the top as they have no registered speaker's population, including created languages such as **Volapük**, **Novial**, or **Interlingua**, extinct language such as **Gothic**, and languages with no registered speakers as **Old English**, **Inupiaq**, **N'ko**, **Tarantino** or **Church Slavonic**.

As for the depth parameter, some languages already mentioned previously appear again, above or close to English :Cree, Kalaallisut, Tigré, Colognian, Dzongkha and Inuktitut. Other can be mentioned, close to English :

LANGUAGE	COUNTRY	SPEAKERS	STATUS
Bhojpuri	India	53M	Stable
Tigrinya	Eritrea	9.9M	Wide
Lak	Russia	0.1M	Stable
Kalmyk	Russia	0.5M	Endangered
Doteli	Nepal	0.5M	Stable
Sango	Central African Rep.	5.2M	Wide

For WikiBooks, the indigenous languages which are mentioned are: **Bashkir** and **Chuvash**, from Russia; **Shan**, from Myanmar and **Minangkabau**, endangered from Indonesia

For WikiQuotes, the indigenous languages which are mentioned are **Igbo** from Nigeria; **Gun**, from Benin; **Banjar**, from Indonesia; **Central Bikol** from Philippines; **Gorontalo**, endangered language from Indonesia.

For WikiSource, let's highlight the many languages from India which are present, among the 81 languages: **Tamil**, **Telugu**, **Hindi**, **Sanskrit**, **Marathi**, **Kannada**, **Malayalam**, **Assamese**, **Gujarati**, **Punjabi**, **Odia** and **Tulu**.

For Wiktionary, the following indigenous languages are present:

Table 20: Indigenous languages in Wiktionary

LANGUAGE	COUNTRY	STATUS
Mon	Myanmar	Stable
Saraiki	Pakistan	Stable
Shan	Myanmar	Wide
Kabardian	Russia	Endangered
Oromo	Ethiopia	Stable
Dimli (Zazaki)	Turkey	Endangered
Nahuatl	Mexico	Large family
Sinhala	Sri-Lanka	Official
Venetian	Italy	Stable
Upper Sorbian	Germany	Endangered
Batak Mandailing	Indonesia	Endangered
Kara-Kalpak	Uzbekistan	Stable
Shawiya (Tachawit)	Algeria	Stable
Goan Konkani	India	Stable
Tyap	Nigeria	Stable
Māori	New Zealand	Endangered
Tulu	India	Stable
Nauru	Nauru	Official
Manx	Isle of Man	Dormant

After Wikipedia, Wiktionary is the most multilingual of the elements of Wikimedia which allows several endangered languages to store part of their dictionary.

Table 21: Number of languages per Wikimedia element sorted

WIKIMEDIA	NUMBER OF LANGUAGES
Wikipedia	343
Wiktionary	174
WikiSource	81
WikiBooks	78
WikiQuote	74
WikiNews	31
WikiVersity	18
WikiVoyage	18

CONCLUSION

Wikimedia is the most multilingual realm in the Internet and offers reliable and updated statistics on languages. It is a place for major languages, with the lead of English, German and French. However, it is also a place of choice for many indigenous languages including some endangered ones. The internationalization of Wikimedia is an objective which is progressively reached and which deserve careful attention as it could be an interesting thermometer of the growing multilingualism of the Web, however not a proportionate representation of the respective proportion of languages in the Web.

REFERENCES

Pimienta D., Blanco A., Muller de Oliveira, G., (2023), The method behind the unprecedented production of indicators of the presence of languages in the Internet. *Frontiers Research Metrics & Analytics*, Volume 8 - doi: 10.3389/frma.2023.1149347