

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

Observatorio Observatoire Observatory
ESPAÑOL FRANÇAIS ENGLISH



Projeto
“Organização da Presença da Língua Portuguesa na Internet”

Informe final: 31 de agosto de 2021

Dr. Daniel Pimienta e Dr. Gilvan Müller de Oliveira (coordenador)

ÍNDICE

INTRODUÇÃO	3
ENTRADAS DE MODELO	4
SAÍDAS DO MODELO.....	5
VALOR DO MODELO PARA CAPACIDADE, GRADIENTE E INTERNAUTAS	10
ANÁLISE DE RESULTADOS.....	12
ESPECÍFICOS DO PORTUGUÊS.....	13
TENDÊNCIAS GERAIS OBSERVADAS.....	15
CONCLUSÕES E RECOMENDAÇÕES	15
REFERÊNCIAS	17
ANEXO 1: ANÁLISE DE FORÇA DA LÍNGUA	18
ANEXO 2: PORTUGUÊS COMO CANDIDATO L2	22
ANEXO 3: EVOLUÇÃO DA CONECTIVIDADE À INTERNET NOS PAÍSES DE LÍNGUA PORTUGUESA	24
ANEXO 4: ANÁLISE DE CONTEÚDO PROFISSIONAL EXISTENTE E POTENCIAL	25
ANEXO 5: BIBLIOGRAFIA.....	28

TABELAS E FIGURAS

Tabela 1: Língua portuguesa, dados por país	4
Tabela 2: Indicadores de português na Internet (resultados brutos do modelo)	7
Tabela 3: Indicadores para o português.....	7
Tabela 4: Correção de viés para português.....	8
Tabela 5: Português na Internet (resultados corrigidos de vieses).....	9
Tabela 6: Parâmetros para línguas na 6ª posição.....	9
Tabela 7: Perspectivas para línguas na 6ª posição.....	10
Tabela 8: Línguas classificadas por porcentagem de pessoas conectadas.....	10
Tabela 9: Línguas classificadas por capacidade	11
Tabela 10: Línguas classificadas por gradiente	12
Tabela 11: Diferenciação temática ordenada.....	14
Tabela 12: Número de países com falantes	18
Tabela 13: Relações L1 + L2 / L2 mais altas	19
Tabela 14: Indicador de força do idioma.....	20
Tabela 15: Indicador de peso para a força do idioma de hoje.....	20
Tabela 16: Pontuação das línguas recomendadas para L2 do British Council.....	22
Tabela 17: Evolução da conectividade à Internet nos países de língua portuguesa.....	24
Tabela 18: Classificação das línguas majoritárias em Creative Commons	25
Tabela 19: Número de cursos disponíveis por língua.....	26
Tabela 20: Número de artigos no SCIELO por língua.....	26
Tabela 21: Explicação dos diferentes elementos da Wikimédia	27
Tabela 22: Classificação e porcentagem de diferentes elementos da Wikimédia	27

INTRODUÇÃO

Este documento apresenta os resultados de um estudo realizado entre fevereiro de 2021 e agosto de 2021 sobre a presença da Língua Portuguesa na Internet, comparada com outras línguas. O estudo é financiado pelo [Departamento Cultural e Educacional do Ministério das Relações Exteriores do Brasil](#), dentro da moldura do [Instituto Internacional da Língua Portuguesa \(IILP\)](#), da [Comunidade dos Países de Língua Portuguesa \(CPLP\)](#) e executado pelo [Observatório da Diversidade Linguística e Cultural na Internet](#), sob a coordenação da [Cátedra UNESCO em Políticas Linguísticas para o Multilinguismo](#), com sede na Universidade Federal de Santa Catarina (UFSC), Brasil.

A medição dos indicadores da Língua Portuguesa na Internet baseia-se na utilização do modelo desenhado em 2017 pelo Observatório da Diversidade Linguística e Cultural na Internet¹. A metodologia associada a este modelo é descrita em detalhes na referência [2] ou, de forma resumida, na referência [1]. O modelo foi aprimorado em muitos aspectos nesta nova edição e todas as fontes foram atualizadas; as mudanças metodológicas e a lista de fontes são documentadas em um documento separado acessível na referência [3].

Este relatório dedicar-se-á aos resultados do modelo, tratando do português em comparação com o resto das línguas; a apresentação e discussão metodológica, incluindo a análise de vieses é deixada como referência [3].

O modelo que sustenta o cálculo dos vários indicadores da presença da língua na Internet assenta-se sobre um amplo conjunto de fontes, das quais duas fontes principais devem ser destacadas:

- *Ethnologue Global Dataset # 24* de março de 2021, para dados demolingüísticos;
- ITU, para a porcentagem de pessoas conectadas à Internet por país.

Aviso: *O estudo que produz o modelo referenciado é essencialmente um trabalho estatístico baseado em uma grande variedade de fontes de entrada. A adoção de uma fonte importante neste tipo de trabalhos também implica a adoção lógica das regras que sustentam os dados dessa fonte. Os autores não são, portanto, responsáveis pela lista de países e territórios considerados, estabelecida pela ITU, uma agência das Nações Unidas, ou pelas estatísticas de porcentagem de pessoas conectadas à Internet, ou pela lista de línguas com mais de cinco milhões de falantes L1 de acordo com Ethnologue e nem pelo reagrupamento em macrolínguas adotadas pela Ethnologue, em concordância com a norma ISO 693.3.*

¹ O Observatório (<http://funredes.org/lc>) tem trabalhado com o tema do espaço das línguas na Internet desde 1998 (ver [6]) e anteriormente orientou o estudo especificamente para a língua portuguesa (ver [4] e [5]).

ENTRADAS DE MODELO

A tabela a seguir reúne os dados que alimentam o modelo e calcula o número e a porcentagem de falantes de português conectados por país (L1 e L2²) e a taxa de conexão à Internet de cada país; destes dados calcula-se o número de falantes de português (L1 + L2) ligados à Internet, por país, bem como a porcentagem que este número representa em relação ao total de falantes de português L1 + L2.

Tabela 1: Língua portuguesa, dados por país³

PAÍS	L1	L2	% Conectados	Falantes conectados	% do total conectado
TOTAL	232 420 480	25 226 000	67,16%	173 036 439	100%
Andorra	11 700		95,00%	11 115	0,0%
Angola	12 100 000	6.000.000	16,00%	2 896 000	1,7%
Argentina	58 000		83,00%	48 140	0,0%
Austrália	48 900		96,23%	47 056	0,0%
Áustria	3 200		87,75%	2 808	0,0%
Bélgica	52 300		90,28%	47 216	0,0%
Bermudas	2 270		98,37%	2 233	0,0%
Brasil	201 000 000	10 200 000	73,91%	156 103 073	90,2%
Canadá	222.000		91,00%	202 020	0,1%
Cabo Verde	11 800	356 000	61,94%	227 828	0,1%
China-Macau	3 680	15 000	86,47%	16 152	0,0%
Dinamarca	6 930		98,00%	6 791	0,0%
Timor Leste	600	5 000	30,00%	1 680	0,0%
Guiné Equatorial			28,00%	0	0,0%
Finlândia	3 020		89,61%	2 706	0,0%
França	959 000		83,34%	799 228	0,5%
Guiana Francesa	13 900		Incluído em	França	
Alemanha	185 000		88,13%	163 049	0,1%
Guiné-Bissau	5 450	391 000	5,00%	19 823	0,0%
Índia	250 000		40,00%	100 000	0,1%
Irlanda	4 620		84,52%	3 905	0,0%
Jamaica	6 000		55,07%	3 304	0,0%
Japão	270 000		92,73%	250 372	0,1%
Jersey	4 530		90,00%	4 077	0,0%
Líbano	2 960		78,18%	2 314	0,0%
Luxemburgo	74 600		97,06%	72 408	0,0%
Malawi	18 000		15,00%	2 700	0,0%
Moçambique	5 000 000	8 130 000	21,00%	2 757 300	1,6%
Países Baixos	16 600		93,29%	15 486	0,0%
Nova Zelândia	5 620		93,00%	5 227	0,0%

² L1 significa língua materna (também chamada de primeira língua) e L2 significa segunda língua. Uma pessoa pode ou não ter uma L2 e pode ter várias L2. Pressupõe-se que a pessoa tenha fluência suficiente para se comunicar em uma determinada língua para contá-la como L2.

³ Fontes: ITU, Banco Mundial, Ethnologue e Observatório da Diversidade Linguística e Cultural na Internet

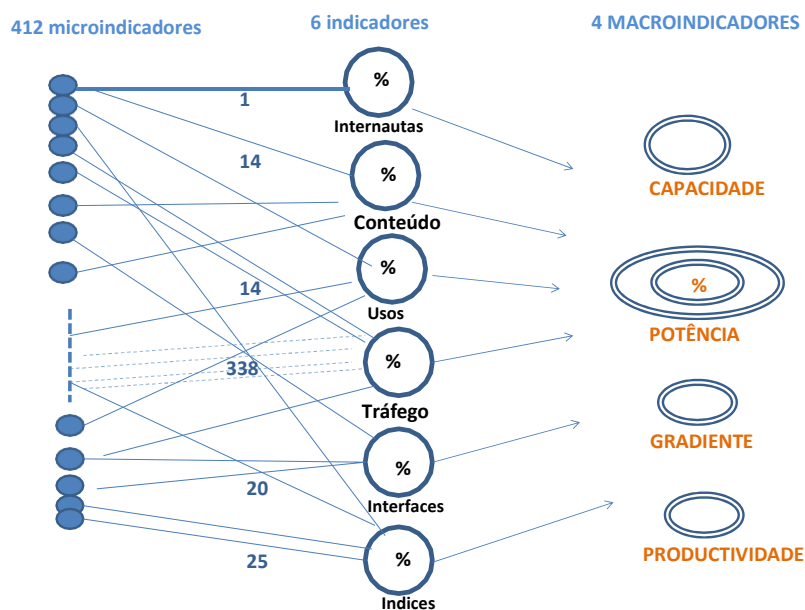
Noruega	10 600		98,00%	10 388	0,0%
Paraguai	235.000		68,52%	161 016	0,1%
Portugal	9 900 000		75,35%	7 459 291	4,3%
Federação Russa	11 200		82,64%	9 256	0,0%
São Tomé e Príncipe	68 000	129 000	33,00%	65 010	0,0%
África do Sul	300 000		60,00%	180 000	0,1%
Espanha	196 000		90,72%	177 809	0,1%
Suécia	14 400		94,54%	13 614	0,0%
Suíça	210 000		93,15%	195 607	0,1%
Reino Unido	135.000		92,52%	124 897	0,1%
Estados Unidos	693.000		88,50%	613 297	0,4%
Uruguai	30 600		76,95%	23 547	0,0%
Venezuela	254.000		72,00%	182 880	0,1%
Zâmbia	5 000		14,30%	715	0,0%
Zimbábue	17 000		30,00%	5 100	0,0%
TOTAL	232 420 480	25 226 000	67,16%	173 036 439	100,0%

Alguns elementos poderiam ser destacados a partir desse quadro, que podem auxiliar na formulação de políticas públicas para a Língua Portuguesa, encontram-se desenvolvidos no Anexo 1.

SAÍDAS DO MODELO

A partir desses dados, o modelo existente fornece os seguintes resultados brutos para os principais macroindicadores sobre a presença de línguas na Internet.

Figura 1: O processo do modelo



Potência: é a média de vários indicadores que medem o espaço de diferentes elementos da Internet (*internautas, índices, interfaces, usos, tráfego, conteúdo*) e deve ponderar a presença global de cada língua na Internet.

Capacidade: é a razão da potência pela porcentagem da população mundial L1 + L2 e, portanto, é um indicador da força de uma língua na Internet, independentemente do seu número de falantes.

Gradiente: é a razão da potência pela porcentagem de falantes L1 + L2 conectados à Internet de uma língua e, portanto, um indicador do dinamismo dos falantes dessa língua conectados (para produzir conteúdo ou usar aplicativos, por exemplo).

Tanto a *capacidade* quanto o *gradiente* são valores sem dimensão e normalizados para 1. A *potência* é uma porcentagem sobre o número mundial de falantes L1 + L2.

Observe que se o total de falantes L1 é igual à população mundial⁴ o total de falantes L1 + L2 é muito maior, pois muitas pessoas falam mais de uma língua. Usando as figuras do Ethnologue, os números correspondentes atuais são:

Total mundial L1 =	7 231 699 136
Total mundial L2 =	3 130 017 620
Total mundial L1 + L2 =	10 361 716 758
Fator de bilinguismo L2 / L1 =	43,28%

Neste documento, todas as porcentagens, salvo indicação em contrário, estão relacionadas com a figura L1 + L2, uma vez que o multilinguismo se aplica a todos os elementos do modelo (o site pode estar desenhado em várias línguas, assim como e-mails da mesma pessoa podem ser escritos em diferentes línguas, o tráfego do usuário pode direcionar para sites em diferentes línguas e assim por diante).

Para mais detalhes sobre os indicadores e uma descrição detalhada da metodologia implementada no modelo, consulte as referências [1] e [3].

Os resultados da língua portuguesa, em comparação com as outras línguas importantes em termos de *potência*, são apresentadas na tabela a seguir, mostrando os resultados para os macroindicadores de *potência*, *capacidade* e *gradiente* e onde as demais colunas devem ser lidas dessa forma:

Internautas: porcentagem mundial de falantes de português L1 + L2 conectados à Internet (em comparação com o total de conectados L1 + L2).

Pop. Mundial: a porcentagem mundial de falantes de português L1 + L2.

Conect.: a porcentagem de falantes de português L1 + L2 conectados à Internet.

Assim, a frase referente à língua portuguesa pode ser lida como “*os falantes de português representam 2,5% da população mundial e 3,1% das pessoas conectadas à Internet no mundo, 67,2% dos falantes de português estão conectados à Internet*”, tendo em conta que todas as porcentagens mundiais nesse estudo não são comparadas com a população mundial (acima de

⁴ Supondo a simplificação de que os seres humanos falam assim que nascem.

7 bilhões), mas sim com a população mundial de falantes L1 + L2 (acima de 10 bilhões), como forma de ter em devida conta o multilinguismo.

Tabela 2: Indicadores de português na Internet (resultados brutos do modelo)

	INTERNAUTAS	POP. MUNDIAL	CONECT.	POTÊNCIA	CAPACIDADE	GRADIENTE
Inglês	15,3%	13,0%	64,3%	26,48%	2,04	1,73
Chinês	17,7%	14,7%	65,6%	13,92%	0,95	0,79
Espanhol	7,0%	5,2%	73,1%	8,73%	1,67	1,25
Francês	3,0%	2,6%	63,7%	3,71%	1,44	1,24
Hindi	4,3%	5,8%	40,2%	3,38%	0,58	0,79
Português	3,1%	2,5%	67,2%	3,35%	1,35	1,10
Russo	3,5%	2,5%	67,2%	3,11%	1,25	0,88
Árabe	3,9%	3,5%	60,1%	3,09%	0,88	0,80
Alemão	2,1%	1,3%	40,2%	2,86%	2,19	1,37
Japonês	2,1%	1,2%	92,6%	2,52%	2,07	1,22
Malaio	2,2%	2,4%	51,0%	1,76%	0,75	0,80
Italiano	0,9%	0,7%	75,7%	1,37%	2,09	1,51
Turco	1,2%	0,9%	78,0%	1,24%	1,46	1,02
Coreano	0,9%	0,8%	64,7%	0,96%	1,22	1,03
Bengali	1,1%	2,6%	81,2%	0,88%	0,34	0,78
RESTO	31,8%	40,4%		22,64%		
TOTAL	100%	100%		100%		

Os outros indicadores da presença da língua portuguesa na Internet (que servem para calcular os 3 macroindicadores anteriores) são os seguintes (repare novamente que essas percentagens se referem ao multilinguismo):

Tabela 3: Indicadores para o português

INDICADOR	%	EXPLICAÇÃO
Internautas	3,05	Os falantes de português representam 3,05% da população mundial conectada à Internet.
Tráfego	1,42	De 100% do tráfego da Web 1,4% vem de falantes de português
Uso	5,53	De 100% dos assinantes das diferentes redes ou plataformas sociais 5,5% são falantes de português
Conteúdo	3,30	Entre todos os sites, 3,3% são escritos em português
Interfaces	3,85	Este é um percentual mundial que modula a % de falantes de português conectados ⁵ desde o número de vezes que o português é usado em interfaces de aplicativos ou como opção na tradução online.
Índices	2,92	Esta é uma porcentagem mundial que modula a % de falantes de português conectados da avaliação do país a que pertencem em termos de uma variedade de parâmetros da Sociedade da Informação (como qualidade de governo eletrônico, por exemplo)

⁵Uma presença elevada, como é o caso das interfaces em português, aumenta o valor acima da percentagem de pessoas ligadas; uma classificação abaixo da média (como nos índices) diminuiria o valor abaixo. Observe que os valores calculados dependem da imagem completa das outras línguas. Os resultados são apresentados dessa forma para permitir total consistência entre os indicadores e permitir o cálculo dos macroindicadores. Para obter explicações detalhadas sobre o que é apenas um procedimento matemático usado para garantir a homogeneidade dos resultados, consulte a referência [1].

O dado mais significativo é a forte presença de falantes de português entre os assinantes das redes sociais, o que implica uma orientação menos profissional para o *conteúdo*. Consulte o Anexo 4 para uma análise do *conteúdo* profissional existente e potencial.

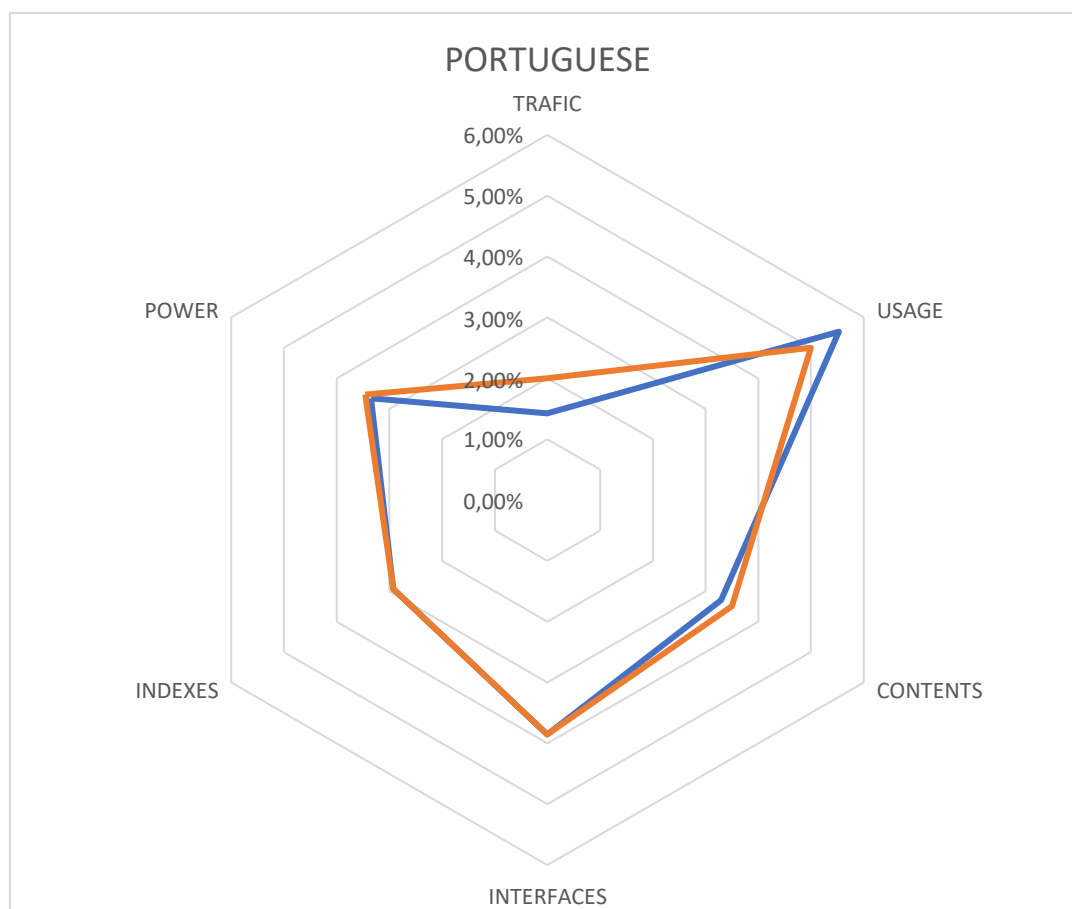
Esses números, calculados a partir de diversas fontes, não estão isentos de vieses; esses vieses são discutidos na referência [1] e o próximo capítulo propõe algum tipo de correção dos vieses.

FIGURAS CORRIGIDAS DE VIESES PARA POTÊNCIA E CONTEÚDO

Diferentes vieses foram identificados durante o processamento das fontes e a implementação do modelo, vieses que podem afetar negativamente os resultados para algumas línguas; isso acontece principalmente para algumas línguas da Ásia, mas também para o indicador de tráfego do português. A correção dos valores do português é apresentada na tabela a seguir.

Tabela 4: Correção de viés para português

PORTUGUÊS	TRÁFEGO	USO	CONTEÚDO	INTERFACES	ÍNDICE	POTÊNCIA
MODELO	1,42%	5,53%	3,30%	3,85%	2,92%	3,35%
CORREÇ. DE VIESES	2,0%	5,5%	3%	3,9%	2,9%	3,5%



A tabela a seguir é uma tentativa de compensar os vieses identificados para as línguas mais poderosas. Verifica-se, após correção do viés, que o português pertence ao quinto grupo de línguas na Internet junto com o russo, um pouco acima do árabe, alemão e japonês.

Tabela 5: Português na Internet (resultados corrigidos de vieses)

LÍNGUA	CONTEÚDO	PRODUTIVIDADE
Inglês	25,00%	1,92
Chinês	15,00%	1,02
Espanhol	7,00%	1,34
Francês	4,00%	1,55
Hindi	4,00%	0,69
Português	3,50%	1,41
Russo	3,50%	1,41
Árabe	2,50%	0,71
Alemão	2,50%	1,92
Japonês	2,50%	2,05
Malaio	1,80%	0,76
Italiano	1,40%	2,14
Turco	1,20%	1,41
Coreano	1,20%	1,53
Bengali	1,20%	0,46
Vietnamita	0,70%	0,94
RESTO	23,00%	0,58

As diferentes dinâmicas demográficas e o progresso da brecha digital dessas línguas determinarão as tendências para o futuro. A proporção de falantes de L2, bem como a dispersão de falantes entre os países, podem interferir no potencial das políticas públicas de fortalecimento da presença de uma língua na Internet. Algumas línguas estão intimamente relacionadas a países com alta conectividade e podem ter atingido seu topo e esperar um progresso mais lento, enquanto outras estão associadas a países com uma forte brecha digital a ser superada, o que também significa altas possibilidades de progresso a longo prazo. As duas tabelas a seguir oferecem o conjunto de parâmetros para as línguas próximas que competem pela posição 6 e uma breve análise das perspectivas.

Tabela 6: Parâmetros para línguas na 6ª posição

	L1 + L2 (Milhão)	L1 + L2 / L1	CONECT. %	Número de países com falantes
Português	257,6	1,11	67,2	44
Árabe	366,2	1,06	60,14	61
Alemão	135,0	1,76	87,7	59
Japonês	126,4	1,00	92,62	24
Russo	258,0	1,68	77,20	51

Tabela 7: Perspectivas para línguas na 6ª posição

	PERSPECTIVAS DE CURTO PRAZO	PERSPECTIVAS DE LONGO PRAZO
Português	O português fica junto com o russo e a capacidade de produzir conteúdo pode fazer a diferença.	Provavelmente manterá o segundo lugar do grupo, exceto se a população de falantes de russo L2 mostrar um crescimento importante.
Árabe	O árabe vai competir com o português e o russo se conseguir reduzir rapidamente a diferença em termos de porcentagem de falantes conectados.	Se tornará o primeiro do grupo quando sua brecha digital maior for superada.
Alemão	Perto do topo, a conectividade e a população L2 mais alta são os principais fatores.	Manterá a posição logo acima do japonês.
Japonês	Alcançou o topo de conectividade. A curto prazo alcançará a última posição do grupo.	Manterá a última posição do grupo sem nenhuma alavanca clara para alterá-la.
Russo	Alguma margem de crescimento com conectividade e falantes L2.	Tendência demográfica mostra crescimento inferior ao português ⁶ .

Essas mesmas cinco línguas estão definitivamente competindo para se tornarem línguas L2 estratégicas em países interessados em fortalecer sua posição em um mundo cada dia mais globalizado e digitalizado (ver Anexo 1 para uma análise desenvolvida pelo *British Council*).

VALOR DO MODELO PARA CAPACIDADE, GRADIENTE E INTERNAUTAS

Observe que os seguintes resultados são a saída bruta do modelo sem correção de viés e que há uma sensibilidade importante para o indicador de conteúdo nos resultados de capacidade e gradiente que tendem a favorecer as línguas com forte presença no reino da Wikimedia⁷.

Tabela 8: Línguas classificadas por porcentagem de pessoas conectadas

LÍNGUA	Internautas
Dinamarquês	97,82%
Sueco	93,49%
Japonês	92,62%
Holandês	92,02%
Alemão suíço	91,56%
Flamengo ocidental	90,43%
Finlandês	89,67%

⁶ A taxa de natalidade atual para o Brasil em 2021 é de 13,313 nascimentos por 1.000 pessoas, um declínio de 1,86% em relação a 2020, enquanto a taxa de natalidade para a Rússia em 2020 foi de 12,194 nascimentos por 1.000 pessoas, um declínio de 2,31% em relação a 2019.

⁷ Por exemplo, o fato de o malgaxe ter uma presença extremamente desproporcional na Wikimedia em comparação com seu número de falantes e sua conectividade muito baixa explicam sua presença na parte superior do indicador de gradiente.

Bávaro	87,68%
Alemão	87,65%
Hebraico	85,46%
Eslovaco	82,47%
Bielo-russo	82,27%
Tcheco	81,37%
Polonês	81,17%
Húngaro	79,92%
Tártaro	78,05%
Turco	77,98%
Servo-croata	77,78%
Grego	77,71%
Russo	77,20%
Cazaque	76,98%
Romeno	75,66%
Italiano	75,65%
Albanês	75,48%
Azerbaijani	74,76%
Napoletano-calabrese	74,39%
Espanhol	73,08%
Curdo	73,02%
Búlgaro	70,34%
Armênio	69,86%
Vietnamita	69,04%
Guarani	68,83%
Português	67,16%

É evidente que existe uma brecha digital forte na Lusofonia e muitos países lusófonos precisam aumentar a sua presença na Internet, o que irá desencadear naturalmente um crescimento do conteúdo em língua portuguesa.

Tabela 9: Línguas classificadas por capacidade

LÍNGUA	Capacidade
Hebraico	5,24
Finlandês	3,42
Servo-croata	3,14
Sueco	2,61
Holandês	2,26
Alemão	2,19
Dinamarquês	2,19
Italiano	2,09
Japonês	2,07
Inglês	2,04
Polonês	1,88
Húngaro	1,79
Grego	1,75
Tcheco	1,70
Espanhol	1,67
Turco	1,46

Francês	1,44
Armênio	1,41
Português	1,35
Eslovaco	1,30
Russo	1,25

Tabela 10: Línguas classificadas por gradiente

LÍNGUA	Gradiente
Hebraico	3,35
<i>Servo-croata</i>	2,21
<i>Malgaxe</i>	2,21
Finlandês	2,09
Inglês	1,73
Sueco	1,53
Italiano	1,51
Alemão	1,37
Holandês	1,34
Polonês	1,26
Espanhol	1,25
Francês	1,24
Grego	1,23
Dinamarquês	1,22
Húngaro	1,22
Japonês	1,22
Tcheco	1,14
Armênio	1,11
Português	1,10

A classificação dos indicadores de *capacidade* e *gradiente* mostra que há muito espaço para os falantes de português avançarem na aplicação das políticas de produção de conteúdo. Como um lembrete, o *gradiente* deve refletir o comportamento dos usuários na Internet (a intensidade de seu tráfego, o número de assinaturas em redes sociais, com quantos artigos da Wikipedia eles contribuíram, etc.).

O indicador de *capacidade* destaca claramente os países que são conhecidos por seu impulso na Sociedade da Informação e cuja população de falantes está concentrada em países desenvolvidos (ou simplesmente no mesmo país). O português tem muito espaço para melhorias ao tentar transformar seus falantes mais em produtores do que em meros consumidores.

ANÁLISE DE RESULTADOS

A figura dos falantes de português ligados à Internet poderá ser realçada atendendo aos seguintes fatos, por ordem de importância:

- Dados extremamente baixos da Internet em Guiné-Bissau e Angola;
- Dados muito baixos da Internet em Moçambique;

- Portugal apresenta números de Internet relativamente baixos em comparação com outros países da OCDE e baixa taxa de crescimento (ver Anexo 3 para mais detalhes).

Observe-se que a população de falantes de português mais importante de pessoas conectadas por país pertence, respectivamente, ao Brasil (90%⁸), Portugal (4,3%⁹), Angola (1,7%), Moçambique (1,6%), França (0,5%) e Estados Unidos (0,4%).

Um modelo de simulação com um cenário de aumento da penetração da Internet de Angola, Guiné Bissau e Moçambique para 50% e de Portugal para 85% mostra que a *conectividade* a Internet do português aumentará para 71,5% (em vez de 67,16%) e o seu indicador de *potência* aumentará a 3,47% (em vez de 3,35%), fazendo com que o português cruze ligeiramente o hindi e alcance a 5ª posição.

Outra simulação com um cenário de 85% de pessoas conectadas no Brasil (deixando os dados dos demais países inalterados) produz um efeito mais forte, com *conectividade* de 76,25% (ao invés de 67,16%) e aumento do valor de *potência* para 3,58% (ao invés de 3,35%) provocando a ultrapassagem do hindi pelo português, e deixando o português apenas 0,11% abaixo do francês (nesse cenário a *capacidade* aumentaria logicamente em 1,44 e o *gradiente* diminuiria em 1,04 porque o cenário mantém as outras entradas inalteradas).

Finalmente, uma terceira simulação com os 2 cenários anteriores juntos mostra o português como a quarta língua da Internet muito ligeiramente acima do francês com 3,70% contra 3,71%.

Este exercício demonstra que os dados globais do espaço do português na Internet são de fato muito dependente dos dados brasileiros, mas que uma política coordenada para enfrentar a brecha digital em todos os países de língua portuguesa produziria efeitos positivos, ainda mais se fosse organizada em conjunto com uma sólida política de produção de conteúdo em português.

ESPECÍFICOS DO PORTUGUÊS

O modelo permite, manipulando estatísticas de microindicadores por país, quando transformados em indicadores de língua, observar, para cada tema, as diferenças significativas entre os resultados temáticos médios e a média geral, a fim de determinar os pontos fortes e fracos em cada tema sob análise. Esta análise baseia-se essencialmente na observação do tráfego para os países lusófonos num conjunto de sites assinalados como refletindo uma determinada temática.

Sendo a amostragem estatística bastante reduzida pelo enfoque temático, em relação ao total, os resultados devem então ser tomados com cautela; no entanto, isto mostra a tendência da natureza do tráfego em português e permite algum tipo de classificação do mesmo. É assim possível perceber se os internautas lusófonos utilizam a Internet mais pelo viés profissional ou do lazer (não profissional). A tabela é ordenada a partir dos temas em que a diferença percentual em relação à média é positiva e maior em direção aos temas em que a diferença é negativa e maior.

⁸ Enquanto o Brasil representa 82% do total da população mundial de língua portuguesa L1 + L2 calculada a partir dos números do Ethnologue.

⁹ Enquanto Portugal representa 3,8% da população lusófona mundial L1 + L2, calculada a partir dos números do Ethnologue.

Foram observados os seguintes resultados para o português, por ordem de importância, sendo em cor azul as temáticas mais visitadas, em preto aquelas sem diferença significativa e em vermelho as temáticas com um claro déficit de tráfego em relação à média geral.

Tabela 11: Diferenciação temática ordenada

TEMAS	ORIENTAÇÃO profissional vs lazer¹⁰
Motores de busca	
Jogos	lazer
Filme	lazer
Hospedagem na web	profissional
Pornô	lazer
Compartilhamento de arquivos	
Ciência e Tecnologia / Pesquisa	profissional
Redes sociais orientadas para a música	lazer
Ferramentas do computador	profissional
Correio eletrônico	profissional
Redes sociais voltadas para imagens	lazer
Redes sociais de namoro	lazer
Serviços de mensagens	lazer
Aplicativos orientados para marketing	profissional
Livro	
Redes sociais de amizade	
Vídeo	lazer
Portais	lazer
MOOC	profissional
Formulários	
ICT	Profissional
Redes sociais profissionais	Profissional
Resposta da questão (Ex. Quora)	
Blog	Profissional
comércio eletrônico	Profissional

Esta análise tende a confirmar a impressão de que os falantes de português são mais consumidores do que produtores de conteúdo e que a utilização da Internet tem mais peso para lazer do que para o aspecto profissional.

A lista de temas está momentaneamente reduzida ao que é mostrado na tabela; na futura expansão do modelo, prevê-se estender tanto o número de sites analisados quanto ao tráfego, quanto o número de temas diferenciados, a fim de se obter uma análise mais apurada.

¹⁰ Nenhuma menção é feita quando o tema poderia estar dos dois lados.

TENDÊNCIAS GERAIS OBSERVADAS

A observação da evolução dos resultados desde 2017 mostra as seguintes tendências:

- ✓ A presença relativa do inglês na Internet continua diminuindo, juntamente com a ampliação da internacionalização da Internet. O inglês estava perto de 30% em 2017; agora está perto de 25%.
- ✓ O chinês agora tem mais falantes conectados à Internet do que o inglês, embora seu conteúdo e aplicativos fiquem cerca de 50% mais baixos do que para o inglês.
- ✓ O espanhol consolida-se como terceira língua da Internet.
- ✓ Como os países francófonos africanos demoram mais para superar a brecha digital do que o crescimento médio da Internet, a quarta posição dos franceses se enfraquece.
- ✓ O árabe pode ser uma das línguas que mais cresce junto com o turco.
- ✓ O japonês pode ter atingido seu máximo e, portanto, experimenta um ligeiro declínio relativo, sendo substituído por árabe e hindi no grupo que ocupa a quinta posição.

CONCLUSÕES E RECOMENDAÇÕES

A língua portuguesa divide na Internet a sexta posição junto com um conjunto de outras 4 línguas: árabe, russo, alemão, japonês com pesos próximos hoje. A perspectiva de curto prazo é que a língua portuguesa possa liderar este grupo; a longo prazo, o árabe deve levar vantagem demográfica e o português deve se consolidar no 7º lugar. Existem vários espaços e possibilidades para a língua portuguesa reforçar a sua presença no ciberespaço a partir de políticas públicas adequadas.

A CPLP pode considerar uma prioridade estratégica o reforço da presença do português na Internet, visto que as apostas são não só culturais e diplomáticas, mas também socioeconômicas. Se for desenhado um esforço coordenado entre os países da CPLP, deve-se considerar as seguintes ações como prioritárias:

- Apoiar a luta contra a brecha digital nos países africanos de língua oficial portuguesa e apoiar a capacitação do Instituto de Estatística de cada país para incluir ou ampliar em inquéritos nacionais questões sobre pessoas ligadas à Internet. Propor um papel de expertise local a Cabo Verde para transferência de know-how e experiência para o resto do continente africano, dada a sua posição de destaque relativo.
- Desenvolver programas de *Letramento Midiático e Informacional*, com especial enfoque na criação de conteúdo, juntamente com a melhoria da conectividade à Internet, em todos os países da CPLP. Desenvolver parcerias com a UNESCO / IFAP, a Cátedra UNESCO de Políticas Linguísticas para o Multilinguismo e outros órgãos para ajudar a

alcançar este objetivo. Reforçar a presença do português em conteúdo aberto (Wikimedia, Creative Commons, MOOCs, ...).

- Reforçar o desenvolvimento de programas de promoção do português como candidato à aprendizagem como língua estrangeira, visando a ampliação estratégica dos falantes como L2.
- Desenvolver ações de sensibilização (cursos, seminários, materiais de informação) dirigidas ao corpo profissional diplomático dos países da CPLP sobre a importância de políticas de fomento à presença da língua portuguesa, seus falantes e culturas na Internet.
- Buscar ideias e propor medidas concretas para associar as línguas indígenas/nativas/nacionais dos países da CPLP aos programas mencionados, o que dá destaque ao português como língua de acesso e abertura.

Para finalizar o relatório, propõe-se um esquema de valorização dos pontos **Fortes, Oportunidades, Fracos e Ameaças** (FOFA) da língua portuguesa no ciberespaço, como síntese das constatações e orientação para o desenvolvimento de políticas públicas adequadas de apoio à presença da língua no ciberespaço.

Figura 2: Esquema FOFA



Nota: O tema das indústrias da música e sua importância na globalização mundial deve merecer um estudo específico¹¹.

¹¹ Veja por exemplo: *A Mobilidade Como Eixo Estruturante da CPLP Para os Próximos Anos, Para Potenciar a Cultura e as Indústrias Criativas*, Vitor Da Sousa.

https://repositorium.sdum.uminho.pt/bitstream/1822/73731/1/2021_Sousa_A-Mobilidade-Como-Eixo-Estruturante-da-CPLP.pdf

REFERÊNCIAS

- [1] D. Pimienta, Indicators of Languages in the Internet, *LT4ALL International Conference Language Technologies for All : Enabling Linguistic Diversity and Multilingualism Worldwide*, UNESCO., Paris, November 2019.
<http://funredes.org/lc2017/Indicators%20Language%20Internet.pdf>
- [2] D. Pimienta, “Uma abordagem alternativa para produzir indicadores de línguas na Internet”, Observatório da Diversidade Linguística e Cultural na Internet, junho de 2017
<http://funredes.org/lc2017/Alternativa%20Lingua%20Internet.docx>
- [3] D. Pimienta, “Versão nova e aprimorada de uma abordagem alternativa para a produção de indicadores da presença de línguas na Internet”, agosto de 2021
<http://funredes.org/lc2021>
- [4] D. Pimienta, “A Língua Portuguesa na Internet e no Mundo Digital”, Entrevista ao IILP, 2012 - <http://iilp.wordpress.com/2012/03/23/entrevista-com-daniel-pimienta/>
- [5] D. Pimienta, “El espacio de las lenguas en la Internet: panorama y tendencias con enfoque especial hacia el Portugués”, in Proc. do terceiro colóquio do IILP “A Língua Portuguesa na Internet e no Mundo Digital”, IILP, Fortaleza, 2012
<http://funredes.org/presentation/IILP-Fortaleza%20Portugues.ppt>
- [6] D. Pimienta, D. Prado and Á. Blanco, “Twelve years of measuring linguistic diversity on the Internet: balance and perspectives”, UNESCO, CI-2009/WS/1
<http://unesdoc.unesco.org/ulis/cgi-bin/ulis.pl?catno=187016>
- [7] British Council, “Languages for the Future (Which languages the UK needs most and why)”, 2013
<https://www.britishcouncil.org/research-policy-insight/policy-reports/languages-future-2017>
- [8] Valnora Leister, editor, “Brazil and the Emergence of a Digital Lusosphere”, Lexington Books, 2018 – ISBN 978149855074
- [9] Luis Reto, Fernando Luís Machado José Paulo Esperança, “Novo atlas da língua portuguesa”, Imprensa Nacional Casa da Moeda, 2018 - ISBN-10 9722726072
- [10] António Branco, “A língua portuguesa na era digital”, Coleção Livros Brancos, METANET, Springer - 2012
<http://www.meta-net.eu/whitepapers/e-book/portuguese.pdf>
- [11] Gilvan Müller de Oliveira, “Política linguística e internacionalização: a língua portuguesa no mundo globalizado do século XXI, *Trabalhos em linguística aplicada* 52, 409-433 - 2013
- [12] Gilvan Müller de Oliveira, “Um Atlântico ampliado: o português nas políticas linguísticas do século XXI”, *O português no século XXI: Cenário geopolítico e sociolinguístico*. São Paulo, 2013

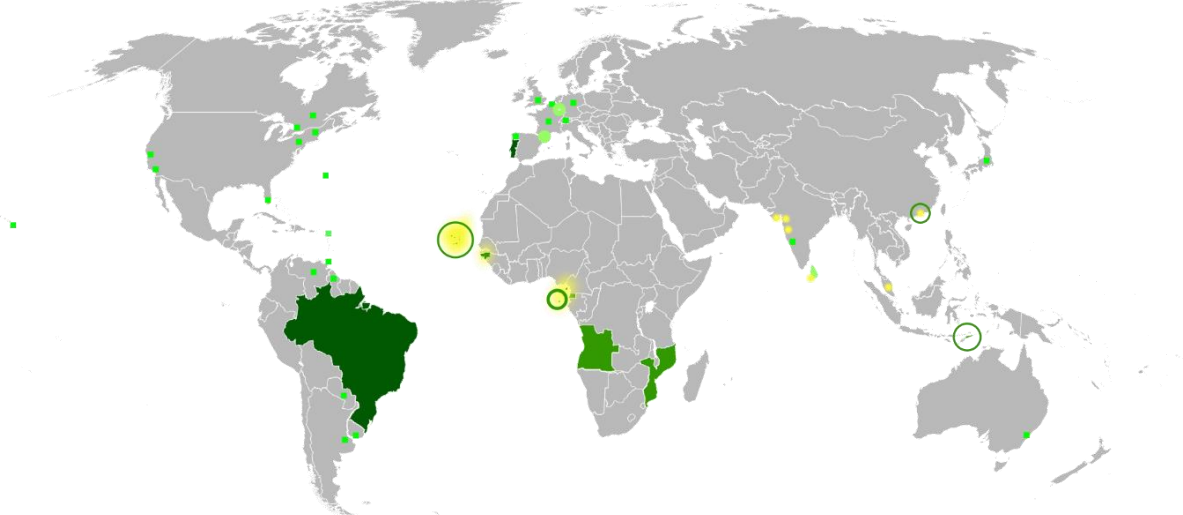
ANEXO 1: ANÁLISE DE FORÇA DA LÍNGUA

O número de países com falantes de português é de 44, o que coloca o português na 10ª posição nesse critério, entre as 138 línguas com mais de 5 milhões de falantes, embora longe dos cinco primeiros.

Tabela 12: Número de países com falantes

LINGUA	Número de países com falantes
Inglês	148
Francês	108
Espanhol	70
Árabe	61
Alemão	59
Russo	51
Chinês	47
Italiano	46
Português	44
Grego	41
Turco	39
Ucraniano	37
Hindi	35
Polonês	34

Figura 3: Mapa mundial da língua portuguesa



Mapa da língua portuguesa no mundo. Verde escuro: Língua nativa. Verde: idioma oficial e administrativo. Verde claro: idioma cultural ou secundário. Amarelo: crioulo de origem portuguesa. Praça verde: minorias que falam português **Fonte: Jonatan Argento - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=15336625>**

O valor $L1+L2/L1$ do português é 1,11 o que o torna muito distante das línguas com números elevados de falantes de L2, na ordem dos 50, não muito longe do espanhol.

Tabela 13: Relações L1 + L2 / L2 mais altas

LÍNGUA	L1 + L2 POP.	NB DE PAÍSES COM ALTO- FALANTES	L2 POP.	L1 + L2 / L2
Jula	12 504 000	3	2 226 000	5,62
<i>Suaíli</i>	80 338 410	22	18 378 410	4,37
Inglês	1 348 145 040	148	369 935 120	3,64
Bamanankan	14 181 800	6	4 181 800	3,39
Francês	266 959 240	108	79 579 970	3,35
Urdu	230 052 270	27	69 006 470	3,33
<i>Malaio</i>	244 761 296	24	81 592 296	3,00
Tailandês	60 715 470	17	20 715 470	2,93
Afrikaans	17 631 780	13	7 331 780	2,40
Sotho, sul	13 524 700	4	5 624 700	2,40
Setswana	13 665 730	4	5 815 730	2,35
Xhosa	19 183 300	3	8 183 300	2,34
Zulu	27 770 100	6	12 070 100	2,30
Wolof	12 266 290	4	5 926 290	2,07
Ganda	11 003 450	3	5 613 450	1,96
Amárico	57 445 260	13	32 345 260	1,78
Alemão	134 993 040	59	76 540 740	1,76
Ibibio	10 439 000	2	5 939 000	1,76
Hindi	600 484 970	35	342 218 070	1,75
.....				
Bengali	267 653 920	24	228 651 040	1,17
.....				
Espanhol	542 894 260	70	471 397 120	1,15
<i>Chinês</i>	1 525 335 340	47	1 325 506 940	1,15
.....				
Finlandês	5 770 040	13	5 114 040	1,13
.....				
Português	257 646 480	44	232 420 480	1,11
.....				
Turco	88 100 870	39	82 230 570	1,07
Holandês	24 382 600	20	22 777 600	1,07

No entanto, um indicador mais significativo integraria ambos os fatores, a proporção L1+L2/L1 e o número de países onde a língua é falada, já que uma alta proporção L2 em um número limitado de países não representa a mesma força como ocorre quando o número de países é alto. O indicador é construído multiplicando a proporção L1+L2/L1 pela porcentagem de países do mundo onde a língua é falada. Com essa métrica, indicadora de potencialidade, o português sobe para a posição 13, mais perto das línguas com as quais concorre no mundo globalizado.

Tabela 14: Indicador de força do idioma

LÍNGUA	L1 + L2 / L1	NB DE PAÍSES COM ALTO-FALANTES	INDICADOR	% INDICADOR
Inglês	3,64	148	2,23	16,87%
Francês	3,35	108	1,50	11,33%
Alemão	1,76	59	0,43	3,25%
<i>Suaíli</i>	4,37	22	0,40	3,01%
Urdu	3,33	27	0,37	2,82%
Russo	1,68	51	0,35	2,68%
Espanhol	1,15	70	0,33	2,52%
<i>Malaio</i>	3,00	24	0,30	2,25%
<i>Árabe</i>	1,06	61	0,27	2,01%
Hindi	1,75	35	0,25	1,92%
<i>Chinês</i>	1,15	47	0,22	1,69%
Tailandês	2,93	17	0,21	1,56%
Português	1,11	44	0,20	1,53%
Italiano	1,05	46	0,20	1,51%
Ucraniano	1,21	37	0,19	1,40%
<i>Persa</i>	1,27	33	0,17	1,31%
Turco	1,07	39	0,17	1,31%
Grego	1,00	41	0,17	1,29%
Polonês	1,02	34	0,14	1,08%

Finalmente, um último indicador pondera o anterior com a porcentagem de falantes conectados à Internet para mostrar a força real ao invés do potencial.

Tabela 15: Indicador de peso para a força do idioma de hoje

LÍNGUA	L1 + L2 / L1	NB PAÍSES	% L1 + L2 CONN.	INDIC. ML	INDIC. ML%	INDIC. PONDERADO.	INDIC. PONDERADO. %
Inglês	3.64427	148	64,33%	2,23	16,87%	1,43	18,43%
Francês	3.35460	108	63,67%	1,50	11,33%	0,95	12,26%
Alemão	1.76368	59	87,65%	0,43	3,25%	0,38	4,85%
Russo	1.67895	51	77,20%	0,35	2,68%	0,27	3,51%
Espanhol	1.15167	70	73,08%	0,33	2,52%	0,24	3,13%
<i>Árabe</i>	1.05560	61	60,14%	0,27	2,01%	0,16	2,06%
<i>Malaio</i>	2.99981	24	51,00%	0,30	2,25%	0,15	1,95%
Italiano	1.04754	46	75,65%	0,20	1,51%	0,15	1,94%
<i>Chinês</i>	1.15076	47	65,59%	0,22	1,69%	0,15	1,88%
Tailandês	2.93092	17	66,85%	0,21	1,56%	0,14	1,77%
Português	1,10854	44	67,16%	0,20	1,53%	0,14	1,74%
Turco	1.07139	39	77,98%	0,17	1,31%	0,13	1,73%
Grego	1.00443	41	77,71%	0,17	1,29%	0,13	1,70%
Ucraniano	1.21185	37	63,96%	0,19	1,40%	0,12	1,52%
Polonês	1.01663	34	81,17%	0,14	1,08%	0,12	1,49%
<i>Persa</i>	1,27234	33	64,58%	0,17	1,31%	0,11	1,44%
Romena	1.00000	34	75,66%	0,14	1,06%	0,11	1,37%
Hindi	1.75469	35	40,18%	0,25	1,92%	0,10	1,31%
<i>Servo-croata.</i>	1.06730	29	77,78%	0,13	0,97%	0,10	1,28%

Japonês	1.00096	24	92,62%	0,10	0,75%	0,09	1,18%
Suaíli	4,37135	22	22,84%	0,40	3,01%	0,09	1,17%
Urdu	3,33378	27	24,12%	0,37	2,82%	0,09	1,15%
Tcheco	1,25028	21	81,37%	0,11	0,82%	0,09	1,13%
Curdo	1.00000	29	73,02%	0,12	0,91%	0,09	1,13%
Coreano	1.00000	32	64,73%	0,13	1,00%	0,09	1,10%

Nota: Sobre a influência global da língua ver o interessante estudo de Bruno Gonçalves¹².

¹² *Links that speak: The global language network and its association with global fame*, Shahar Ronen, Bruno Gonçalves and ale. PNAS 2014 111 (52) E5616-E5622 - <https://doi.org/10.1073/pnas.1410931111>

ANEXO 2: PORTUGUÊS COMO CANDIDATO L2

Hoje, a língua portuguesa tem uma proporção muito baixa de falantes de L2, poucos computados em países industrializados. Existe a possibilidade de nos apoiarmos na força do português na Internet para mudar essa situação? Em um relatório de 2013, o British Council analisa as línguas estrangeiras que o cidadão do Reino Unido deveria aprender prioritariamente (referência [6]). O método utilizado pelo relatório, bem como os resultados, é interessante para trazer alguma luz aos desafios da língua portuguesa como candidata L2 nos próximos anos.

Os indicadores usados para orientar as recomendações são econômicos (1 a 4), culturais, educacionais ou diplomáticos (5 a 8) ou para a última categoria os chamados fatores de equilíbrio (9 e 10):

1. O comércio de exportação atual
2. As necessidades linguísticas do setor empresarial
3. As prioridades comerciais do governo
4. Os mercados emergentes de alto crescimento
5. As prioridades diplomáticas e de segurança
6. Os interesses linguísticos do público
7. Os destinos externos dos viajantes
8. As prioridades da Estratégia de Educação Internacional do governo
9. Os níveis de proficiência em inglês em outros países
10. A prevalência de diferentes línguas na Internet.

Da análise destes 10 indicadores a recomendação geral do relatório conduz à seguinte classificação, com o português na 6ª posição:

Tabela 16: Pontuação das línguas recomendadas para L2 do British Council

LÍNGUA	PONTUAÇÃO
Espanhol	76
Árabe	54
Francês	47
Chinês	45,5
Alemão	43,5
Português	41
Italiano	22,5
Russo e Turco	19
Japonês	17

A análise está centrada no contexto e nas necessidades específicas do Reino Unido, e o Brexit pode ter alterado alguns dos parâmetros. Apesar do fato de que 8 anos se passaram desde a avaliação, e que a Internet é bastante diferente agora em relação ao que era em 2013, há algumas lições interessantes a tirar dessa análise, no que diz respeito à projeção da língua portuguesa num mundo cotidiano mais globalizado e digitalizado.

Os elementos para a classificação do português foram os seguintes:

- ✓ O português não é uma das onze línguas principais dos atuais mercados de exportação mais importantes do Reino Unido (nenhum país de língua portuguesa nos 15 principais países de exportação).
- ✓ O português ocupa o 8º lugar num inquérito às empresas britânicas que avaliam as línguas consideradas úteis para a sua organização.
- ✓ Porém, o Brasil está na lista dos países considerados como mercado de alto crescimento e finalmente obtém a posição 6 (atrás do holandês) no ranking de línguas necessárias para fins económicos.
- ✓ O português está em terceiro lugar, antes do francês, e muito próximo do árabe, para fins culturais, educacionais e diplomáticos.
- ✓ A proficiência em inglês é muito baixa no Brasil e moderada em Portugal (surgindo a necessidade de compensação).
- ✓ Os portugueses ocupavam o 5º lugar em termos de utilizadores da Internet.

Não é por acaso que as mesmas línguas que competem com o português pela quinta posição na Internet (árabe, alemão, russo e japonês) se encontram na mesma tabela.

ANEXO 3: EVOLUÇÃO DA CONECTIVIDADE À INTERNET NOS PAÍSES DE LÍNGUA PORTUGUESA

É importante saber que a UIT desenvolvia suas próprias estimativas nos países onde não existem fontes oficiais para esse valor. No entanto, a ITU parou de fazer isso a partir de 2018. Como consequência, onde não existe nenhuma fonte oficial ou do Banco Mundial, o Observatório definiu um valor de projeção linear aproximada para permitir que o modelo funcione sem penalização desses países.

Tabela 17: Evolução da conectividade à Internet nos países de língua portuguesa

PAÍS	2016	2017	2018	2019	2019 dado por	2019/2017
Angola	13,00	14,34		16,00	Observatório	
Brasil	60,87	67,47	70,43	73,91	Fonte oficial	8,7%
Cabo Verde	50,32	57,16		61,94	Fonte oficial	7,7%
China-Macau	81,64	83,17	83,79	86,47	Fonte oficial	3,8%
Timor Leste	25,25	27,49		30,00	Observatório	
Guiné Equatorial	23,78	26,24		28,00	Observatório	
Guiné-Bissau	3,76	3,93		5,00	Observatório	
Moçambique	17,52	20,77		21,00	Banco Mundial	1,1%
Portugal	70,42	73,79	74,66	75,35	Fonte oficial	2,1%
São Tomé e Príncipe	28,00	29,93		33,00	Observatório	

A taxa de crescimento do Brasil e Cabo Verde é adequada, não tanto a de Portugal. Já para os demais países, a recomendação é começar a monitorar oficialmente esse importante dado para poder acompanhar as políticas públicas contra a brecha digital.

ANEXO 4: ANÁLISE DE CONTEÚDO PROFISSIONAL EXISTENTE E POTENCIAL

Os principais resultados parecem indicar um uso generalizado da Internet em português que se inclina mais para o lazer e merece ser reforçado do lado dos usos e conteúdo profissionais. No entanto, esse ponto merece atenção mais detalhada antes de formular recomendações.

Na evolução da Internet, a Web 2.0 trouxe uma grande implicação dos internautas e ao mesmo tempo uma reorientação de uma Web 1.0, com uma vocação pensada em termos de conhecimento universal compartilhado, para uma Web 2.0 que, por sua natureza, está ligada às redes sociais, expressas mais em termos de *egomação*¹³.

É claro que quando se definem políticas públicas de promoção de conteúdo na Internet em apoio a uma determinada língua, isso deve ser pensado em termos de conteúdo aberto e conhecimento universal, e não em termos de conteúdo pessoal ou privado.

Quais são os indicadores que permitem avaliar a presença do português nesta linha estratégica? A análise da presença da língua portuguesa em conteúdo aberto, na galáxia da Wikimédia, em MOOCs disponíveis e numa ferramenta de pesquisa de artigos científicos como a SCIELO¹⁴ pode fornecer algumas respostas.

Tabela 18: Classificação das línguas majoritárias em Creative Commons

	2015	2017
Inglês	1	1
Espanhol	3	2
Português	4	3
Alemão		4
Francês	3	5
Coreano	2	

Fonte: <https://stateof.creativecommons.org>

¹³ Neologismo introduzido pelo autor deste relatório (fonte: *Vade retro Internet! Uma visão crítica da evolução da Internet da sociedade civil*, D. Pimienta, L.G. Rodríguez, Revista Ibero-Americana de Ciência da Informação, V.13N.3-2020, pag. 979-1000, <https://periodicos.unb.br/index.php/RICI/article/view/33041/27497>) para qualificar informações fortemente ligadas ao ego do usuário e de que o interesse geral está ausente, exceto, é claro, para as empresas de marketing que as veem como um meio extraordinário de direcionamento de seus anúncios e os apoiadores das redes sociais que fazem dela a essência seu modelo comercial.

¹⁴ SCIELO é uma base de dados bibliográfica, biblioteca digital e modelo de publicação eletrônica cooperativa de periódicos de acesso aberto. A SCIELO foi projetado para atender às necessidades da região e de forma mais geral do Sul Global e tem se beneficiado do apoio da FAPESP, CNPq e BIREME.

O português foi em 2017 a terceira língua mais utilizada no Creative Commons, uma das associações carro-chefe no esforço de dados abertos¹⁵.

No que diz respeito aos MOOCs, um setor de alcance planetário enorme, uma vez que no que diz respeito à educação, o português parece ter progredido entre 2017 e 2021 do nono lugar, na melhor das hipóteses, para o quarto lugar.

Tabela 19: Número de cursos disponíveis por língua

LÍNGUA	NÚMERO DE CURSOS em 2021	% 2021	% 2018
Inglês	36051	81,8%	81,3%
Espanhol	3138	7,1%	8,1%
Francês	1052	2,4%	5,1%
Português	682	1,5%	<0,7%
Russo	621	1,4%	2,6%
Turco	483	1,1%	
Árabe	381	0,9%	1,4%
Chinês	362	0,8%	2,7%
Italiano	289	0,7%	2,2%
Japonês	289	0,7%	
Alemão	251	0,6%	0,7%
Tailandês	224	0,5%	
Hindi	128	0,3%	
OUTROS (18)	120	0,3%	

Fonte: <https://www.classcentral.com/languages>

No que diz respeito ao SCIELO, as estatísticas permitem-nos verificar uma forte presença do português, com um certo reequilíbrio ao longo do tempo.

Tabela 20: Número de artigos no SCIELO por língua

	COMPLETO	2010 <DOC <2020	DOC > 2020
Inglês	220450 49,7%	30125 73,0%	152933 61,1%
Português	275769 62,2%	19362 46,9%	138864 55,5%
Espanhol	14188 3,2%	1719 4,2%	9792 3,9%

Fonte: <https://analytics.scielo.org>

Por fim, uma análise completa do lugar do português na galáxia da Wikimédia, em comparação com outras línguas ocidentais importantes, mostra uma presença insuficiente (19ª posição em média) e suscetível de aumentar, em particular pelo número de artigos da Wikipédia e sua atualização, bem como o dicionário aberto de português. A primeira tabela explica os diferentes conceitos abordados pela Wikimédia, sempre com a abertura dos dados como ponto de ancoragem, e a segunda tabela apresenta uma elaboração feita a partir de dados estatísticos da Wikimédia, comparando as respectivas posições do português, inglês, espanhol e alemão. Em conclusão, o português tem ativos suficientes no espaço aberto da Internet e pode visar um contributo de qualidade para o conhecimento universal a partir do reforço das suas posições.

¹⁵ Infelizmente, essas estatísticas não são atualizadas desde 2017.

Tabela 21: Explicação dos diferentes elementos da Wikipédia

ELEMENTO ¹⁶	SIGNIFICADO	EXPLICAÇÃO
WikiP-Form	Wikipédia: fórmula	Uma fórmula foi desenvolvida para integrar os vários parâmetros de presença na Wikipédia.
WikiP-A	Wikipédia: artigos	Número de itens existentes
WikiP-D	Wikipédia: profundidade	A profundidade é um parâmetro definido pela Wikipédia que permite quantificar o grau de atualização dos artigos.
WikiP-U	Wikipédia: editores	Número de pessoas que estão editando artigos ativamente.
WikiP-E	Wikipédia: edições	Número de edições realizadas em artigos, indicador de dinamismo
WikiB	Wikimédia: livros	Número de livros abertos e gratuitos acessíveis
WikiQ	Wikimédia: citações	Número de citações abertas acessíveis
WikiS	Wikimédia: fontes	Quantidade de textos abertos e livres na biblioteca
WikiY	Wikimédia: universidades	Número de recursos educacionais abertos acessíveis
WikiT	Wikimédia: dicionário	Número de entradas no dicionário aberto
WikiN	Wikimédia: notícias	Quantidade de informação jornalística acessível de um ponto de vista neutro
WikiV	Wikimédia: viagens	Número de guias turísticos abertos e gratuitos acessíveis
MED.		Uma classificação e uma percentagem média são estabelecidas a partir da ponderação dos vários elementos.

Tabela 22: Classificação e porcentagem de diferentes elementos da Wikipédia

	WikiP-F	WikiP-A	WikiP-D	WikiP-U	WikiP-E	Wiki-B	Wiki-Q	Wiki-S	Wiki-Y	Wiki-T	Wiki-N	Wiki-V	MED.
Clas. Português	22	18	25	7	11	4	8	15	6	97	18	9	18,71
%	0,2%	1,9%		3,1%	2,1%	4,4%	3,1%	0,8%	3,5%	0,0%	1,0%	3,5%	1,7%
Clas. inglês	1	1	1	1	1	1	1	2	2	1	1	1	1,10
%	55,6%	11,8%		41,8%	34,6%	35,7%	14,3%	23,6%	29,0%	19,2%	14,9%	28,1%	41,1%
Clas. Francês	5	6	12	3	3	3	11	1	3	3	2	5	4,29
% FR	3,0%	4,1%		6,0%	6,1%	6,1%	2,5%	27,9%	13,2%	11,0%	14,3%	7,2%	6,5%
Clas. Espanhol	21	10	17	4	4	8	9	35	8	7	7	11	15,57
%	0,3%	3,0%		5,2%	4,6%	3,7%	3,0%	0,2%	2,0%	2,9%	2,9%	2,5%	1,7%
Clas. Alemão	4	4	56	2	2	2	7	7	1	8	19	2	4,48
%	5,9%	4,5%		6,0%	7,1%	8,9%	3,3%	4,6%	36,9%	2,4%	1,0%	16,1%	8,0%

¹⁶ As seguintes fontes foram usadas para desenvolver esta tabela:

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

<https://meta.wikimedia.org/wiki/Wikibooks/Table>

<https://stats.wikimedia.org/wikiquote/FR/Sitemap.htm>

<https://meta.wikimedia.org/wiki/Wikisource>

<https://stats.wikimedia.org/wikiversity/EN/Sitemap.htm>

<https://stats.wikimedia.org/wiktionary/EN/Sitemap.htm>

<https://stats.wikimedia.org/wikinews/EN/Sitemap.htm>

<https://stats.wikimedia.org/wikivoyage/EN/Sitemap.htm>*

ANEXO 5: BIBLIOGRAFIA

A seguir, é proposta uma lista de referências em torno do tema da presença da língua portuguesa na Internet.

ANO	TÍTULO	AUTOR	MODELO	Link
2018	O Brasil e o surgimento de uma lusofonia digital	Valnora Leister	Livro	1
2018	Novo atlas da língua portuguesa	Reto, Luís.	Artigo	2
2012	A Língua Portuguesa na Era Digital (METANET)	António Branco	Livro	3
2016	Língua portuguesa global: um estudo de caso luso-brasileiro	José Gabriel Andrade	Conferência	4
2006	Impacto Económico da Língua Portuguesa Enquanto Língua de Trabalho	Maria Souza Galito	Tese	5
2014	Corpora na Linguatca: Visão e caminhos percorridos	Diana Santos	Capítulo	6
2017	O valor económico da língua portuguesa: um olhar geral	Reto, Luís.	Artigo	7
2012	Potencial Económico da Língua Portuguesa	José Paulo Esperança	Artigo	8
2014	A comunidade dos países de língua portuguesa: o papel da língua num mundo globalizado	Edalina Rodrigues Sanches	Artigo	9
2014	Política da Língua e Globalização: A Língua Portuguesa no Século XXI	Gilvan Müller De Oliveira	Capítulo	10
2016	Caracterizando a Mineração de Conteúdo de Texto Gerado pelo Usuário: Um Estudo Sistemático de Mapeamento da Língua Portuguesa	Ellen Souza	Artigo	11
2008	Uma abordagem eclética para a avaliação da língua: a influência global da língua portuguesa *	José Paulo Esperança	Artigo	12
2014	Política de línguas no espaço virtual: língua brasileira e (ou) língua portuguesa (?)	Machado, Júlio Cesar	Artigo	13
2015	A identidade da língua portuguesa no mundo: aventuras e desventuras de uma língua internacional	Diana Santos	Capítulo	14
2007	Existe ciência além do inglês?	Abel Packer	Artigo	15
2009	Análise e desenho do portal de webquests em língua portuguesa	Bottentuit Junior	Artigo	16
2012	Um olhar crítico sobre a sociometria da língua portuguesa	Diego barbosa da silva	Artigo	17
2009	Uma Abordagem Eclética ao Valor da Língua: O Uso Global do Português	José Paulo Esperança	Artigo	18
2017	A lexicografia portuguesa na era da Internet	Teresa Lino	Capítulo	19
2002	Medindo a Web em português	Rachel Aires, Diana Santos	Artigo	20
2006	Práticas de multilinguismo e linguagem comercial na Internet	Helen Kelly-Holmes	Artigo	21
2019	Memes e educação da Internet: uma sequência didática para as aulas de história e língua portuguesa	Maria Alice de Souza	Artigo	22
2016	Caracterizando a Mineração de Opinião: Um Estudo Sistemático de Mapeamento da Língua Portuguesa	Ellen Souza	Artigo	23
2004	Recursos de acesso público (gratuito) da Internet para a língua portuguesa	Edward A. Riedinger	Artigo	24
2012	Alfabetização de surdos: um processo computacional para projetar artefatos em língua de sinais / português para a Internet	Camila Souza de Oliveira Guimarães	Conferência	25
2013	Análise de Sentimentos no Twitter para a Língua Portuguesa	Eduardo Santos Duarte	Tese	26

2012	Análise de sentimento em dados do Twitter para a língua portuguesa	Marlo Souza	Artigo	27
2014	BrasilEduca - Uma plataforma MOOC de código aberto para falantes de português com conceitos de gamificação	Marceli AA Mesquita	Artigo	28
2021	Os estudantes chineses e os hábitos de uso da Internet ao encontro da língua portuguesa e do mundo lusófono	Manuel João Pires	Artigo	29
2018	BlogSet-BR: Um Blog Corpus em português do Brasil	Henrique DP dos Santos	Artigo	30
2003	O caso de um motor de pesquisa na web português	Mário J. Silva	Artigo	31
2007	Construção de comunidades internacionais em português	JC Paiva	Artigo	32
2014	O português como recurso comunicativo em um mundo globalizado: o como e o porquê dos novos rumos na construção de teorias	Luiz Paulo Moita-Lopes	Capítulo	33
2019	Desenvolvimento de Sistema para Classificação Automática de Notícias Falsas em Língua Portuguesa	Roger Oliveira Monteiro	Artigo	34
2018	Culturas orais e multilinguismo em um mundo de big data: O caso dos países de língua portuguesa	Claudia Wanderley	Artigo	35
2004	Diversidade cultural e direito à comunicação	Tadao Takahashi	Artigo	36
2005	Caracterizando uma Rede Comunitária Nacional	Daniel Gomes, Mario J. Silva	Artigo	37

LISTA DE URLs CORRESPONDENTES	
1	https://books.google.com/books?id=nfBIDwAAQBAJ
2	https://www.researchgate.net/publication/324258714_Novo_Atlas_da_Lingua_PortuguesaNew_Atlas_of_the_Portuguese_Language_sample
3	http://www.meta-net.eu/whitepapers/e-book/portuguese.pdf
4	http://www.lasics.uminho.pt/ojs/index.php/cecs_ebooks/article/download/2643/2552
5	https://www.researchgate.net/profile/Maria-Sousa-Galito/publication/264159387_Impacto_Economico_da_Lingua_Portuguesa_Inquerito_Internacional_Distribuido_em_Paises_Lusofonos_PALOP/links/53d058f90cf25dc05cfe3d0b/Impacto-Economico-da-Lingua-Portuguesa-Inquerito-Internacional-Distribuido-em-Paises-Lusofonos-PALOP.pdf
6	https://www.linguateca.pt/Diana/download/SantosCorporaLinguateca2013.pdf
7	https://www.researchgate.net/publication/323364665_O_valor_economico_da_lingua_portuguesa_um_olhar_geral
8	https://www.academia.edu/33920958/Potencial_Económico_da_Língua_Portuguesa
9	https://www.academia.edu/download/39473422/The_community_of_Portuguese_Language_Spe20151027-7454-113pexd.pdf
10	https://www.taylorfrancis.com/chapters/edit/10.4324/9781315757391-7/language-policy-globalization-portuguese-language-twenty-first-century-gilvan-m%C3%BCller-de-oliveira
11	https://www.researchgate.net/publication/305458190_Characterizing_User-Generated_Text_Content_Mining_A_Systematic_Mapping_Study_of_the_Portuguese_Language
12	http://efnil.org/documents/conference-publications/lisbon-2008/06-EFNIL.Lisbon-publ.Esperanca.pdf
13	https://www.academia.edu/download/35739347/Dialogos-Pertinentes-v.10-n.1.pdf#page=62
14	https://www.linguateca.pt/Diana/download/LNI.pdf
15	https://www.embopress.org/doi/full/10.1038/sj.embor.7400906
16	https://repositorium.sdum.uminho.pt/handle/1822/9829
17	https://www.academia.edu/5286669/Um_olhar_crítico_sobre_a_sociometria_da_língua_portuguesa

18	http://cvc.institutocamoes.pt/conhecer/biblioteca-digital-camoes/lingua-1%252F1228-1228%252Ffile.html
19	https://www.taylorfrancis.com/books/edit/10.4324/9781315104942/routledge-handbook-conceptsicography-pedro-fuertes-olivera
20	https://www.linguateca.pt/Diana/download/posterAiresSantosEuroWeb2002.pdf
21	https://doi.org/10.1111/j.1467-9841.2006.00290.x
22	https://go.gale.com/ps/i.do?id=GALE%7CA581730618&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=19849540&p=IFME&sw=w&userGroupName=anon~95bb512b
23	https://link.springer.com/chapter/10.1007/978-3-319-41552-9_12
24	https://www.scielo.br/j/hcsm/a/B9sMFSVYSPQKGxGYSDf5z4g/?lang=en
25	https://www.researchgate.net/publication/235793354_Deaf_Literacy_A_Computational_Process_to_Design_Sign_LanguagePortuguese_Artifacts_for_Internet
26	https://run.unl.pt/handle/10362/11338
27	https://www.researchgate.net/publication/262175717_Sentiment_Analysis_on_Twitter_Data_for_Portuguese_Language
28	https://www.researchgate.net/publication/274081853_BrasilEduca_-An_Open-Source_MOOC_platform_for_Portuguese_speakers_with_gamification_concepts
29	https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3833821
30	https://aclanthology.org/L18-1105.pdf
31	https://repositorio.ul.pt/bitstream/10451/14125/1/03-3.pdf
32	https://www.academia.edu/10326268/Construction_of_international_communities_in_Portuguese
33	https://www.researchgate.net/publication/289606804_Portuguese_as_a_communicative_resource_in_a_globalized_world_The_how_and_why_of_new_directions_in_theory-_building
34	https://www.researchgate.net/profile/Rodrigo-Nogueira-16/publication/340738742_Development_of_a_System_for_Automatic_Classification_of_Fake_News_in_Portuguese_Language/links/5e9b0bbb14585150839e5d79ake-forews-Classification-Development-Automático-Automático-Sistema-oficial-Autocent-Autuguese--evelopment---pt-ag-in-ofuguese-pt-pt-pt-pt-pt-ofauta-pt-a--pt- -tuguese---evelopment-in-ofugu---pt- -tugues-Auta- tugu-of-ugu-ofugue---evelopment- Language.pdf
35	https://content.iospress.com/articles/education-for-information/efi180210
36	http://www.campus-oei.org/pensariberoamerica/ric06a05.htm
37	http://xldb.fc.ul.pt/daniel/gomesCharacterizing.pdf

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

Observatorio Observatoire Observatory
ESPAÑOL FRANÇAIS ENGLISH



United Nations
Educational, Scientific
and Cultural Organization



UNESCO Chair on
Language Policies for Multilingualism
Federal University of Santa Catarina (UFSC)
Florianópolis, Brazil

ESTUDO SOBRE A PRESENÇA DE LÍNGUA PORTUGUESA NA INTERNET

AGOSTO 2021

Daniel Pimienta, Observatório da Diversidade Lingüística e Cultural na Internet

Gilvan Müller de Oliveira (Coord.), Cátedra UNESCO de Políticas Linguísticas para Multilinguismo

Obrigado:

A Álvaro Blanco por escrever programas que mudaram radicalmente o manuseio de tantas fontes e ortografias de línguas e países, e a David Pimienta, que escreveu as macros Excel necessárias para transformar o formato Ethnologue no formato exigido para este estudo.

➤ O projeto começou em meados de fevereiro de 2021

➤ É financiado por *Departamento Cultural e Educacional do Ministério das Relações Exteriores do Brasil*

➤ dentro da moldura do

➤ É coordenado por

➤ É realizado por



United Nations
Educational, Scientific
and Cultural Organization



• UNESCO Chair on
• Language Policies for Multilingualism
• Federal University of Santa Catarina (UFSC)
• Florianópolis, Brazil

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

Observatorio Observatoire Observatory
ESPAÑOL FRANÇAIS ENGLISH

❑ O projeto está apoiado no **novo método** definido em 2017 pelo Observatório, o **atualizar** de seus parâmetros para 2021 números e alguns importantes **melhorias**.

❑ O modelo produz um **conjunto de macro-indicadores**. para uma série de idiomas (até agora idiomas com L1> 5 M.), a partir de um conjunto de 3 tipos de entradas.

❑ A metodologia é completamente descrita em <http://funredes.org/lc2017>

“Uma abordagem alternativa para produzir indicadores de línguas na Internet.”, D. Pimienta, junho de 2017

❑ O método atualizado esta descrito em <http://funredes.org/lc2021>

“Versão nova e aprimorada de uma abordagem alternativa para a produção de indicadores da presença de línguas na Internet”, D. Pimienta, agosto de 2021

O CONCEITO

- Usar **tanto e tão variados quanto possíveis números de dados** da Internet para aproximar o espaço das línguas.
- Tente alcançar um grande escopo de **aplicações e espaços** da Internet.



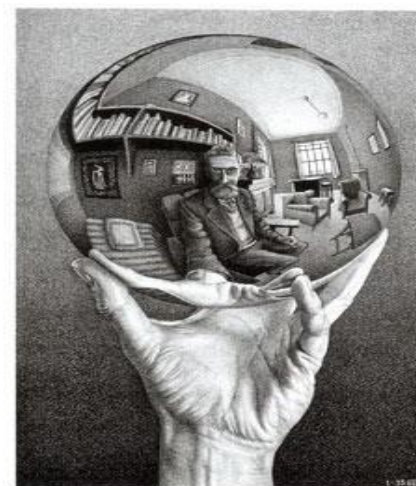
O CONCEITO

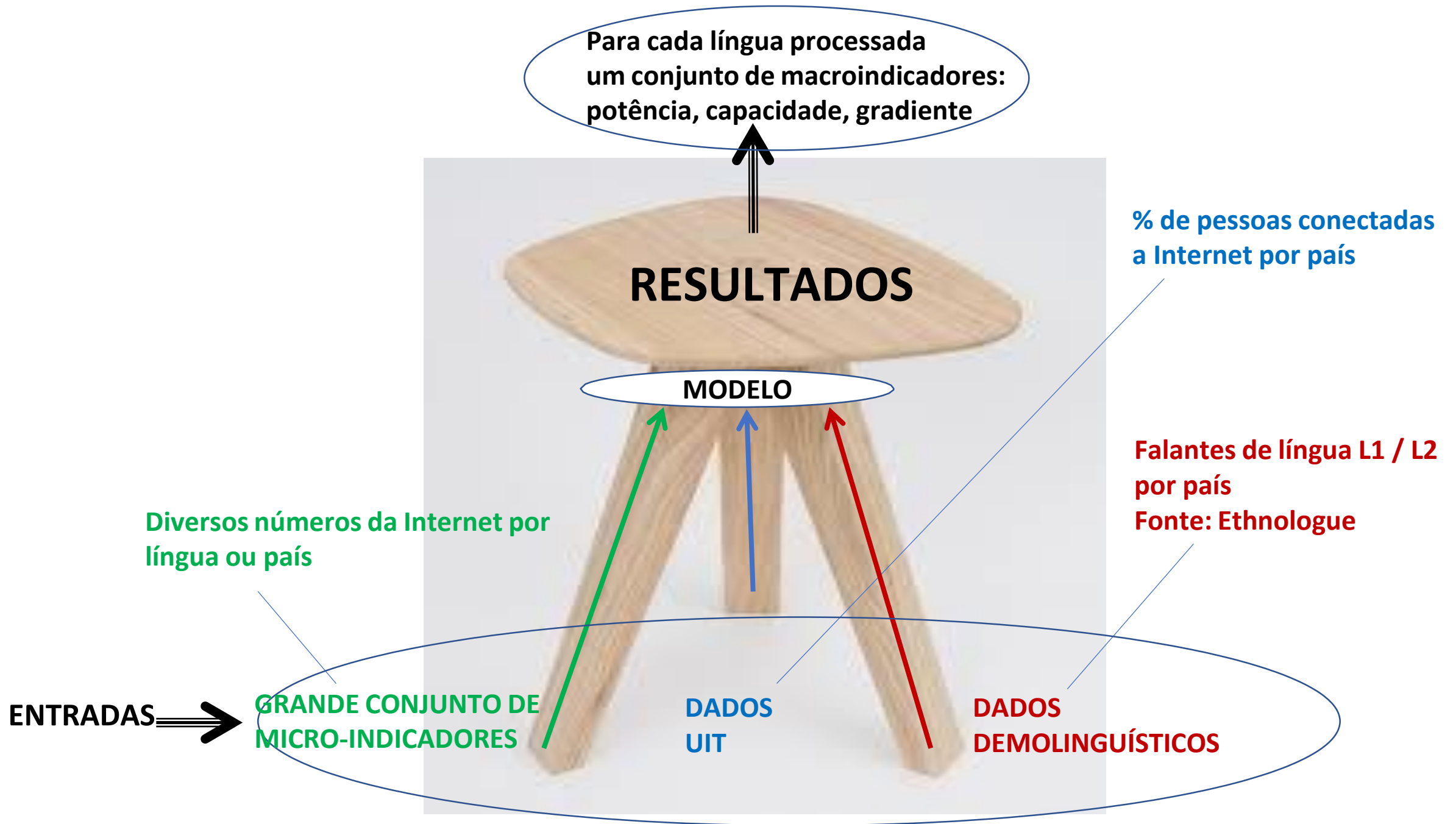
- Figuras linguísticas escassas completas por mais fácil de encontrar **figuras do país**.
- **Transformar** figuras de países em figuras de idiomas.



O CONCEITO

- Use simples **Estatísticas** truques para obter indicadores significativos.
- Preste muita atenção a todos os possíveis **preconceitos**.

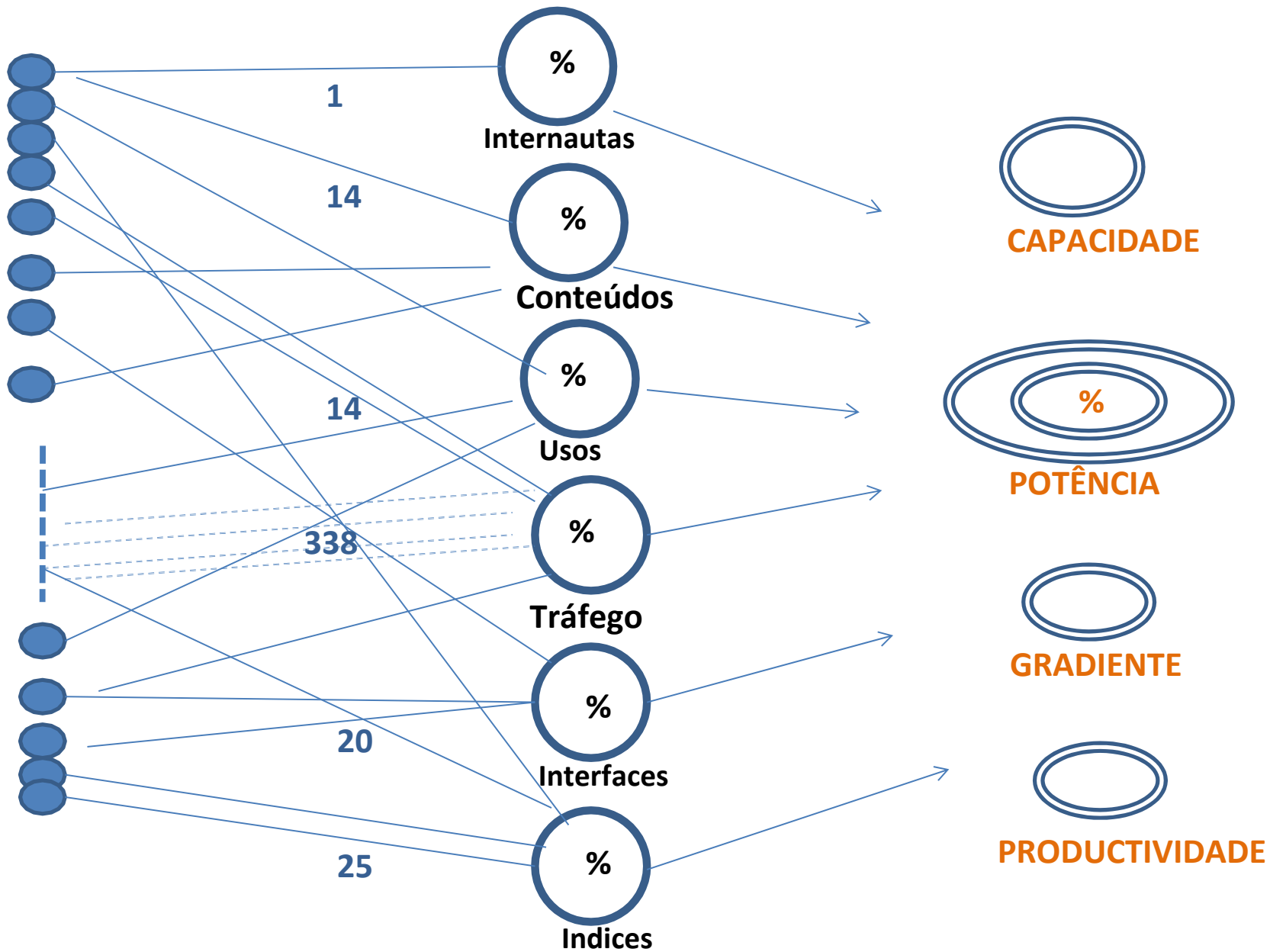




412 microindicadores

6 indicadores

4 MACROINDICADORES



ENTRADAS DO MODELO

PAÍS	L1 + L2	L1	L2	% Conectado	Número	% de total
				por país	Conectado /país	conectado
TOTAL	257 646 480	232 420 480	25 226 000	67,14%	172 979 623	100%
Angola	18 100 000	12 100 000	6.000.000	16,00%	2 896 000	1,7%
Brasil	211 200 000	201 000 000	10 200 000	73,91%	156 103 073	90,2%
Canadá	222.000	222.000		91,00%	202 020	0,1%
Ilhas de cabo verde	367 800	11 800	356 000	61,94%	227 828	0,1%
China-Macau	18 680	3 680	15 000	86,47%	16 152	0,0%
Timor Leste	5 600	600	5 000	30,00%	1 680	0,0%
Guiné Equatorial				28,00%		
França	959 000	959 000		83,34%	799 228	0,5%
Alemanha	185 000	185 000		88,13%	163 049	0,1%
Guiné-bissau	396 450	5 450	391.000	5,00%	19 823	0,0%
Índia	250 000	250 000		40,00%	100 000	0,1%
Japão	270 000	270 000		92,73%	250 372	0,1%
Moçambique	13 130 000	5 000 000	8 130 000	21,00%	2 757 300	1,6%
Paraguai	235.000	235.000		68,52%	161 016	0,1%
Portugal	9 900 000	9 900 000		75,35%	7 459 291	4,3%
São Tomé e Príncipe	197.000	68 000	129 000	33,00%	65 010	0,0%
África do Sul	300 000	300 000		60,00%	180 000	0,1%
Espanha	196 000	196 000		90,72%	177 809	0,1%
Suíça	210 000	210 000		93,15%	195 607	0,1%
Reino Unido	135.000	135.000		92,52%	124 897	0,1%
Estados Unidos	693.000	693.000		88,50%	613 297	0,4%
Venezuela	254.000	254.000		72,00%	182 880	0,1%
OUTROS PAÍSES (22)	421 950	421 950			340 106	0,2%

SAÍDA DE MODELO (RESULTADOS CRUOS)

	INTERNAUTAS	POP. MUNDIAL	CONECT.	POTÊNCIA	CAPACIDADE	GRADIENTE
Inglês	15,3%	13,0%	64,3%	26,48%	2,04	1,73
Chinês	17,7%	14,7%	65,6%	13,92%	0,95	0,79
Espanhol	7,0%	5,2%	73,1%	8,73%	1,67	1,25
Francês	3,0%	2,6%	63,7%	3,71%	1,44	1,24
Hindi	4,3%	5,8%	40,2%	3,38%	0,58	0,79
Português	3,1%	2,5%	67,2%	3,35%	1,35	1,10
Russo	3,5%	2,5%	67,2%	3,11%	1,25	0,88
Árabe	3,9%	3,5%	60,1%	3,09%	0,88	0,80
Alemão	2,1%	1,3%	40,2%	2,86%	2,19	1,37
Japonês	2,1%	1,2%	92,6%	2,52%	2,07	1,22
Malaio	2,2%	2,4%	51,0%	1,76%	0,75	0,80
Italiano	0,9%	0,7%	75,7%	1,37%	2,09	1,51
Turco	1,2%	0,9%	78,0%	1,24%	1,46	1,02
Coreano	0,9%	0,8%	64,7%	0,96%	1,22	1,03
Bengali	1,1%	2,6%	81,2%	0,88%	0,34	0,78
RESTO	31,8%	40,4%		22,64%		
TOTAL	100%	100%		100%		

NÚMEROS CORRIGIDOS DE VIESES

LÍNGUA	CONTEÚDO	PRODUTIVIDADE
Inglês	25,00%	1,92
Chinês	15,00%	1,02
Espanhol	7,00%	1,34
Francês	4,00%	1,55
Hindi	4,00%	0,69
Português	3,50%	1,41
Russo	3,50%	1,41
Árabe	2,50%	0,71
Alemão	2,50%	1,92
Japonês	2,50%	2,05
Malaio	1,80%	0,76
Italiano	1,40%	2,14
Turco	1,20%	1,41
Coreano	1,20%	1,53
Bengali	1,20%	0,46
Vietnamita	0,70%	0,94
RESTO	23,00%	0,58

La língua portuguesa encabeça hoje um grupo de 5 línguas em posição 6.

Os dados demográficos irão desempenhar um papel a longo prazo para fazer diferenças entre essas línguas.

ANALISIS PARAMETROS LINGUAS EN POSICION 6

	L1 + L2 (Milhão)	L1 + L2 / L1	CONNECT.%	Número de países com alto-falantes
Português	257,6	1,11	67,2	44
Árabe	366,2	1,06	60,14	61
Alemão	135,0	1,76	87,7	59
Japonês	126,4	1,00	92,62	24
Russo	258,0	1,68	77,20	51

ANALISIS PERSPECTIVAS LINGUAS EN POSICIAO 6

	PERSPECTIVAS DE CURTO PRAZO	PERSPECTIVAS DE LONGO PRAZO
Português	Vai ficar junto com o russo e a capacidade de produzir conteúdos pode fazer a diferença.	Provavelmente manterá o segundo lugar do grupo, exceto se a população de falantes de russo L2 mostrar um crescimento importante.
Árabe	O árabe vai competir com o português e o russo se conseguir reduzir rapidamente a diferença em termos de porcentagem de falantes conectados.	Se tornará o primeiro do grupo quando sua maior fosso digital for superada.
Alemão	Perto do topo, a conectividade e a população L2 mais alta são os principais fatores.	Manterá a posição logo acima do japonês.
Japonês	Alcançou o topo de conectividade. A curto prazo alcançará a última posição do grupo.	Manterá a última posição do grupo sem nenhuma alavanca clara para alterá-la.
Russo	Alguma margem de crescimento com conectividade e falantes L2.	Tendência demográfica mostra crescimento inferior ao português.

LÍNGUA	INTERNAUTAS
Dinamarquês	97,82%
Sueco	93,49%
Japonês	92,62%
Holandês	92,02%
Alemão, suíço	91,56%
Flamengo ocidental	90,43%
Finlandês	89,67%
Bávaro	87,68%
Alemão	87,65%
Hebraico	85,46%
Eslovaco	82,47%
Bielo-russo	82,27%
Tcheco	81,37%
Polonês	81,17%
Húngaro	79,92%
Tártaro	78,05%
Turco	77,98%
Servo-croata	77,78%
Grego	77,71%
Russo	77,20%
Cazaque	76,98%
Romeno	75,66%
Italiano	75,65%
Albanês	75,48%
Azerbaijani	74,76%
Napoletano-calabrese	74,39%
Espanhol	73,08%
Curdo	73,02%
Búlgaro	70,34%
Armênio	69,86%
Vietnamita	69,04%
Guarani	68,83%
Português	67,16%

SAIDA DE MODELO : OTROS MACROINDICADORES

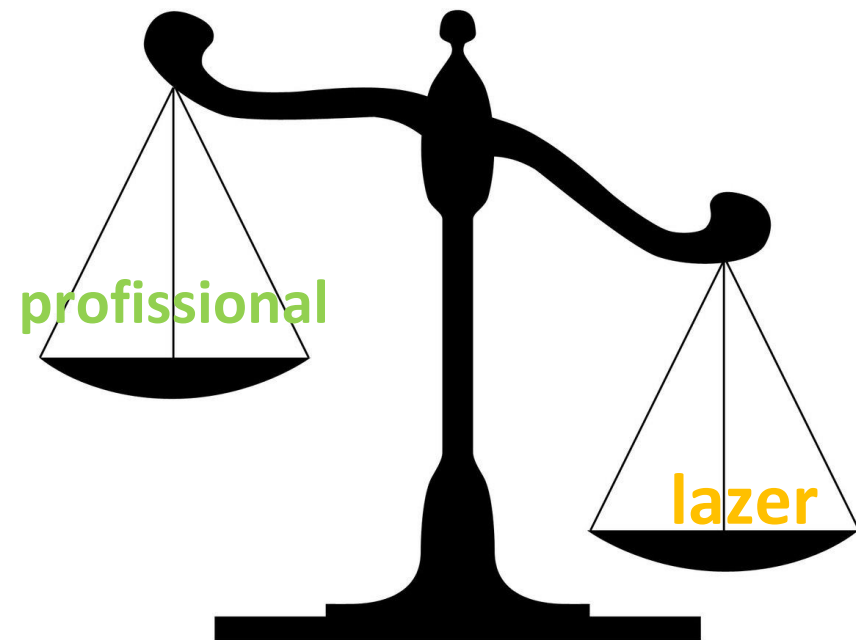
LÍNGUA	CAPACIDADE
Hebraico	5,24
Finlandês	3,42
Servo-croata	3,14
Sueco	2,61
Holandês	2,26
Alemão	2,19
Dinamarquês	2,19
Italiano	2,09
Japonês	2,07
Inglês	2,04
Polonês	1,88
Húngaro	1,79
Grego	1,75
Tcheco	1,70
Espanhol	1,67
Turco	1,46
Francês	1,44
Armênio	1,41
Português	1,35
Eslovaco	1,30
Russo	1,25

LÍNGUA	GRADIENTE
Hebraico	3,35
Servo-croata	2,21
Malgaxe	2,21
Finlandês	2,09
Inglês	1,73
Sueco	1,53
Italiano	1,51
Alemão	1,37
Holandês	1,34
Polonês	1,26
Espanhol	1,25
Francês	1,24
Grego	1,23
Dinamarquês	1,22
Húngaro	1,22
Japonês	1,22
Tcheco	1,14
Armênio	1,11
Português	1,10

TEMAS	ORIENTAÇÃO profissional vs lazer
Motores de busca	
Jogos	lazer
Filme	lazer
Hospedagem na web	profissional
Pornô	lazer
Compartilhamento de arquivos	
Ciência e Tecnologia / Pesquisa	profissional
Redes sociais orientadas para a música	lazer
Ferramentas do computador	profissional
Correio eletrônico	profissional
Redes sociais voltadas para imagens	lazer
Redes sociais de namoro	lazer
Serviços de mensagens	lazer
Aplicativos orientados para marketing	profissional
Livro	
Redes sociais de amizade	
Vídeo	lazer
Portais	lazer
MOOC	profissional
Formulários	
ICT	Profissional
Redes sociais profissionais	Profissional
Resposta da questão	
Blog	Profissional
comércio eletrônico	Profissional



FIGURAS ESPECÍFICAS DO PORTUGUÊS DIFERENCIAÇÃO TEMÁTICA



CONCLUSÃO

- O português pertence a um grupo de 5 línguas que estão em **quinta posição** na internet
- A demografia pode, a longo prazo, colocar os portugueses na **posição 7**
- Falantes de português conectados talvez **mais consumidores do que produtores** e tem mais usos de lazer que profissional então há muita sala para melhorias **capacidade**
- Taxa de conexão para português é impactada por **números muito baixos na África**. No entanto, o **importância demográfica do Brasil** (90% dos conectados) torna-o no local mais sensível para melhoria da presença de português na internet.

RECOMENDAÇÕES

Apoiar a **luta contra a fosso digital** nos países africanos de língua oficial portuguesa

- Apoiar a capacitação do Instituto de Estatística para incluir em **inquéritos nacionais** questões sobre pessoas ligadas à Internet.
 - Propor um papel de **expertise local a Cabo Verde** para transferência de know-how e experiência para o resto do continente, dada a sua posição de destaque relativo.
- Desenvolver programas de *Letramento Midiático e Informacional*, com especial enfoque na criação de conteúdos, juntamente com a melhoria da conectividade à Internet, em todos os países da CPLP.
- Desenvolver parcerias com a UNESCO / IFAP e a Cátedra UNESCO de Políticas Linguísticas para o Multilinguismo para ajudar a alcançar este objetivo.
 - Reforçar a presença do português em conteúdo aberto (Wikimedia, Creative Commons, MOOCs, ...).
- Reforçar o desenvolvimento de programas de promoção da presença do **português como candidato à aprendizagem como língua estrangeira**, visando a ampliação estratégica dos falantes como L2.
- Desenvolver ações de **sensibilização** (cursos, seminários, materiais de informação) dirigidas ao **corpo profissional diplomático dos países da CPLP** sobre a importância de políticas de fomento à presença da língua portuguesa, seus falantes e culturas, na Internet.
- Buscar ideias e propor medidas concretas para associar as **línguas indígenas/nativas/nacionais** dos países da CPLP aos programas mencionados, o que dá destaque ao português como língua de acesso e abertura.

ANALISE FOFA LÍNGUA PORTUGUESA

FORÇA

- 5º idioma em termos de falantes L1
- 44 países com falantes de português
- 9º país em termos de falantes L1 + L2
- Presença mundial dos portugueses nas indústrias culturais, com destaque para a música do Brasil, Portugal e Cabo Verde.

OPORTUNIDADES

- Presença do Brasil na Internet
- Experiência de Internet de Cabo Verde na África
- Potencialidade para crescimento L2
- Vantagem demográfica com concorrentes próximos (exceto árabe)
- Presença do português em conteúdo aberto (Wikimedia, Creative Commons, MOOC...)

FRAQUEZA

- % bastante baixa de falantes conectados
- Forte fosso digital em países africanos de língua portuguesa (exceto Cabo Verde)
- Conectividade de Internet não tão alta em Portugal
- Mais consumidor do que produtor de conteúdos

AMEAÇAS

- Lentidão da superação do fosso digital
- Falta de políticas públicas voltadas para a linguagem
- Mais dinamismo em línguas concorrentes

A group of people, possibly a crowd or a team, is shown in a blue-tinted, semi-transparent overlay. The people are smiling and appear to be in a celebratory or joyful mood. The text "MUITO OBRIGADO" is centered over the image in a bold, white, sans-serif font.

**MUITO
OBRIGADO**

QUADROS ANEXOS EM APOIO A APRESENTAÇÃO

LINGUA	Número de países com alto-falantes
Inglês	148
Francês	108
Espanhol	70
Árabe	61
Alemão	59
Russo	51
Chinês	47
Italiano	46
Português	44
Grego	41
Turco	39
Ucraniano	37
Hindi	35
Polonês	34

**Número de países com alto-falantes
em principais línguas**

LÍNGUA	L1 + L2 POP.	PAÍSES COM ALTO-FALANTES	L2 POP.	L1 + L2 / L2
Jula	12 504 000	3	2 226 000	5,62
Suaíli	80 338 410	22	18 378 410	4,37
Inglês	1 348 145 040	148	369 935 120	3,64
Bamanankan	14 181 800	6	4 181 800	3,39
Francês	266 959 240	108	79 579 970	3,35
Urdu	230 052 270	27	69 006 470	3,33
Malaio	244 761 296	24	81 592 296	3,00
Tailandês	60 715 470	17	20 715 470	2,93
Afrikaans	17 631 780	13	7 331 780	2,40
Sotho, sul	13 524 700	4	5 624 700	2,40
Setswana	13 665 730	4	5 815 730	2,35
Xhosa	19 183 300	3	8 183 300	2,34
Zulu	27 770 100	6	12 070 100	2,30
Wolof	12 266 290	4	5 926 290	2,07
Ganda	11 003 450	3	5 613 450	1,96
Amárico	57 445 260	13	32 345 260	1,78
Alemão	134 993 040	59	76 540 740	1,76
Ibibio	10 439 000	2	5 939 000	1,76
Hindi	600 484 970	35	342 218 070	1,75
.....				
Bengali	267 653 920	24	228 651 040	1,17
.....				
Espanhol	542 894 260	70	471 397 120	1,15
Chinês	1 525 335 340	47	1 325 506 940	1,15
.....				
Finlandês	5 770 040	13	5 114 040	1,13
.....				
Português	257 646 480	44	232 420 480	1,11

PAÍSES COM IMPORTANTE PROPORÇÃO DE FALANTES L2

LÍNGUA	L1 + L2 / L1	NB DE PAÍSES COM ALTO-FALANTES	INDICADOR	% INDICADOR
Inglês	3,64	148	2,23	16,87%
Francês	3,35	108	1,50	11,33%
Alemão	1,76	59	0,43	3,25%
Suaíli	4,37	22	0,40	3,01%
Urdu	3,33	27	0,37	2,82%
Russo	1,68	51	0,35	2,68%
Espanhol	1,15	70	0,33	2,52%
Malaio	3,00	24	0,30	2,25%
Árabe	1,06	61	0,27	2,01%
Hindi	1,75	35	0,25	1,92%
Chinês	1,15	47	0,22	1,69%
Tailandês	2,93	17	0,21	1,56%
Português	1,11	44	0,20	1,53%
Italiano	1,05	46	0,20	1,51%
Ucraniano	1,21	37	0,19	1,40%
Persa	1,27	33	0,17	1,31%
Turco	1,07	39	0,17	1,31%
Grego	1,00	41	0,17	1,29%
Polonês	1,02	34	0,14	1,08%

INDICADOR COMBINADO

$L1+L2/L1 \times NUM / 270$

LÍNGUA	L1 + L2 / L1	NB PAÍSES	% L1 + L2 CONN.	INDIC. ML	INDIC. ML%	INDIC PONDERADO.	INDIC PONDERADO. %
Inglês	3.64427	148	64,33%	2,23	16,87%	1,43	18,43%
Francês	3.35460	108	63,67%	1,50	11,33%	0,95	12,26%
Alemão	1.76368	59	87,65%	0,43	3,25%	0,38	4,85%
Russo	1.67895	51	77,20%	0,35	2,68%	0,27	3,51%
Espanhol	1,15167	70	73,08%	0,33	2,52%	0,24	3,13%
Árabe	1.05560	61	60,14%	0,27	2,01%	0,16	2,06%
Malaio	2.99981	24	51,00%	0,30	2,25%	0,15	1,95%
Italiano	1.04754	46	75,65%	0,20	1,51%	0,15	1,94%
Chinês	1.15076	47	65,59%	0,22	1,69%	0,15	1,88%
Tailandês	2.93092	17	66,85%	0,21	1,56%	0,14	1,77%
Português	1,10854	44	67,16%	0,20	1,53%	0,14	1,74%
Turco	1.07139	39	77,98%	0,17	1,31%	0,13	1,73%
Grego	1.00443	41	77,71%	0,17	1,29%	0,13	1,70%
Ucraniano	1.21185	37	63,96%	0,19	1,40%	0,12	1,52%
Polonês	1.01663	34	81,17%	0,14	1,08%	0,12	1,49%
Persa	1,27234	33	64,58%	0,17	1,31%	0,11	1,44%
Romena	1.00000	34	75,66%	0,14	1,06%	0,11	1,37%
Hindi	1.75469	35	40,18%	0,25	1,92%	0,10	1,31%
Servo-croata.	1.06730	29	77,78%	0,13	0,97%	0,10	1,28%
Japonês	1.00096	24	92,62%	0,10	0,75%	0,09	1,18%
Suaíli	4,37135	22	22,84%	0,40	3,01%	0,09	1,17%
Urdu	3,33378	27	24,12%	0,37	2,82%	0,09	1,15%
Tcheco	1,25028	21	81,37%	0,11	0,82%	0,09	1,13%
Curdo	1.00000	29	73,02%	0,12	0,91%	0,09	1,13%
Coreano	1.00000	32	64,73%	0,13	1,00%	0,09	1,10%

INDICADOR COMBINADO PONDERADO POR CONNECTIVIDADE DA LINGUA

LÍNGUA	PONTUAÇÃO
Espanhol	76
Árabe	54
Francês	47
Chinês	45,5
Alemão	43,5
Português	41
Italiano	22,5
Russo e Turco	19
Japonês	17

Pontuação de British Council
pra recomendar língua de
aprendizagem L2

EVOLUÇÃO DA CONECTIVIDADE À INTERNET NOS PAÍSES DE LÍNGUA PORTUGUESA

PAÍS	2016	2017	2018	2019	2019 dado por	2019/2017
Angola	13,00	14,34		16,00	Observatório	
Brasil	60,87	67,47	70,43	73,91	Fonte oficial	8,7%
Cabo Verde	50,32	57,16		61,94	Fonte oficial	7,7%
China-Macau	81,64	83,17	83,79	86,47	Fonte oficial	3,8%
Timor Leste	25,25	27,49		30,00	Observatório	
Guiné Equatorial	23,78	26,24		28,00	Observatório	
Guiné-Bissau	3,76	3,93		5,00	Observatório	
Moçambique	17,52	20,77		21,00	Banco Mundial	1,1%
Portugal	70,42	73,79	74,66	75,35	Fonte oficial	2,1%
São Tomé e Príncipe	28,00	29,93		33,00	Observatório	

Classificação das línguas majoritárias em Creative Commons

Fonte: <https://stateof.creativecommons.org>

2

	2015	2017
Inglês	1	1
Espanhol	3	2
Português	4	3
Alemão		4
Francês	3	5
Coreano	2	

MOOC: Número de cursos disponíveis por língua

Fonte: <https://www.classcentral.com/languages>

LÍNGUA	NÚMERO DE CURSOS em 2021	% 2021	% 2018
Inglês	36051	81,8%	81,3%
Espanhol	3138	7,1%	8,1%
Francês	1052	2,4%	5,1%
Português	682	1,5%	<0,7%
Russo	621	1,4%	2,6%
Turco	483	1,1%	
Árabe	381	0,9%	1,4%
Chinês	362	0,8%	2,7%
Italiano	289	0,7%	2,2%
Japonês	289	0,7%	
Alemão	251	0,6%	0,7%
Tailandês	224	0,5%	
Hindi	128	0,3%	
OUTROS (18)	120	0,3%	

Número de artigos no SCIELO por língua

Fonte: <https://analytics.scielo.org>

	COMPLETO		2010 <DOC <2020		DOC > 2020	
Inglês	220450	49,7%	30125	73,0%	152933	61,1%
Português	275769	62,2%	19362	46,9%	138864	55,5%
Espanhol	14188	3,2%	1719	4,2%	9792	3,9%

Classificação e porcentagem de diferentes elementos da Wikimédia

	WikiP-F	WikiP-A	WikiP-D	WikiP-U	WikiP-E	Wiki-B	Wiki-Q	Wiki-S	Wiki-Y	Wiki-T	Wiki-N	Wiki-V	MED.
Clas. Português	22	18	25	7	11	4	8	15	6	97	18	9	18,71
%	0,2%	1,9%		3,1%	2,1%	4,4%	3,1%	0,8%	3,5%	0,0%	1,0%	3,5%	1,7%
Clas. inglês	1	1	1	1	1	1	1	2	2	1	1	1	1,10
%	55,6%	11,8%		41,8%	34,6%	35,7%	14,3%	23,6%	29,0%	19,2%	14,9%	28,1%	41,1%
Clas. Francês	5	6	12	3	3	3	11	1	3	3	2	5	4,29
% FR	3,0%	4,1%		6,0%	6,1%	6,1%	2,5%	27,9%	13,2%	11,0%	14,3%	7,2%	6,5%
Clas. Espanhol	21	10	17	4	4	8	9	35	8	7	7	11	15,57
%	0,3%	3,0%		5,2%	4,6%	3,7%	3,0%	0,2%	2,0%	2,9%	2,9%	2,5%	1,7%
Clas. Alemão	4	4	56	2	2	2	7	7	1	8	19	2	4,48
%	5,9%	4,5%		6,0%	7,1%	8,9%	3,3%	4,6%	36,9%	2,4%	1,0%	16,1%	8,0%

Versão nova e aprimorada de uma abordagem alternativa para a produção de indicadores da presença de línguas na Internet

Daniel Pimienta

**Observatório da diversidade linguística e cultural
na Internet**

<http://funredes.org/lc>

Agosto de 2021

Créditos: Esta versão foi possível graças ao apoio do [Departamento de Cultura e Educação do Ministério das Relações Exteriores do Brasil](#) dentro do enquadramento de [Instituto Internacional de Língua Portuguesa](#) e sob a coordenação do [Cátedra UNESCO em Políticas Linguísticas para Multilinguismo](#). Os créditos também são concedidos a Daniel Prado, que foi o primeiro a ter a ideia de coletar fontes múltiplas para medir a presença de línguas na Internet, bem como transformar dados por país em dados por língua.

Obrigado: Ao Professor Gilvan Müller de Oliveira pelo apoio em questões linguísticas e coordenação junto aos financiadores; a Álvaro Blanco por escrever delicadas macros Excel que mudaram radicalmente o manuseio de tantas fontes e ortografias de línguas e países, e a David Pimienta, que escreveu as macros Excel necessárias para transformar o formato Ethnologue no formato exigido para este estudo, bem como para o tratamento de macrolínguas.

Aviso: O estudo a seguir é essencialmente um trabalho estatístico baseado em uma ampla variedade de fontes. A adoção de uma fonte importante neste tipo de trabalho também e logicamente implica a adoção das regras que suportam os dados dessa fonte. O autor não se responsabiliza pela lista de países e territórios considerados, estabelecida pela ITU, órgão das Nações Unidas, nem pela lista de línguas com mais de cinco milhões de falantes L1, segundo o Ethnologue, bem como o agrupamento em macrolínguas, adotado pela Ethnologue, de acordo com a norma ISO 693.3.

RESUMO

Num contexto de escassez de dados fiáveis sobre o espaço linguístico da Internet, foi enriquecida e atualizada a abordagem alternativa de 2017 para a produção de indicadores de comportamento na Internet de 140 línguas com mais de 5 milhões de falantes. São apresentadas melhorias a esta abordagem, que se baseia na compilação de uma grande série de microindicadores sobre línguas ou países em vários espaços ou aplicações na Internet (ou em relação à Internet). A utilização dos últimos resultados produzidos pelo Ethnologue permitiu ter os dados demolinguísticos mais confiáveis e atualizadas e, além disso, fornecer os elementos para superar um dos principais vieses do método, ligado ao tratamento de falantes L2. Os seis indicadores de presença de línguas na Internet definidos e estudados em 2017 (*internautas, tráfego, usos, conteúdo, índices e interfaces*), e os quatro macroindicadores que são deduzidos deles (*potência, capacidade, gradiente e produtividade de conteúdo*) são reproduzidos após as atualizações de todas as fontes para 202. Os resultados mostram uma diminuição relativa da presença do inglês, para uma cifra em torno de 25% (ante 30% em 2017) e o chinês em forte aumento; enquanto o espanhol se consolida na terceira posição. O francês passa a dividir o terceiro lugar com o hindi, com uma pequena vantagem em relação a 2017, em um grupo de línguas com posições muito semelhantes: português, russo, árabe e alemão. Como na edição de 2017, todos os possíveis vieses derivados do método, hipótese ou fontes são examinados e é proposta uma estimativa que leva em conta esses vieses para as principais línguas.

Palavras chave: Línguas, Internet, diversidade linguística, indicadores, vieses

Conteúdo

RESUMO	2
CONTEXTO.....	5
1. INTRODUÇÃO	6
2. DIFERENÇAS DA PRIMEIRA VERSÃO	7
2.1 Adoção do conjunto de dados Ethnologue 24 como uma fonte demolingüística	7
2.2 Gestão de L2 e multilinguismo.....	8
2.3 Fonte para pessoas conectadas à Internet	9
2.4 Gestão de fontes para microindicadores	10
2.4.1 ÍNDICE.....	11
2.4.2 CONTEÚDO	11
2.4.3 TRÁFEGO	13
2.4.4 INTERFACES	14
2.4.5 USOS	15
2.5 Resumo dos indicadores	15
3. RESULTADOS	16
4. ANÁLISE DOS RESULTADOS	19
5. ANÁLISE DE VIESES	21
5.1 Vieses específicos do método	21
5.2 Vieses de seleção de fontes	22
5.3 Vieses das fontes.....	22
5.3.1 Vieses da Wikimédia.....	24
5.3.2 Vieses da Alexa.....	31
5.4 Correção de vieses	32
6. CONCLUSÕES E PERSPECTIVAS	36
REFERÊNCIAS	37
ANEXO 1. LISTA DE MICROINDICADORES E FONTES	38
ANEXO 2: MACROLÍNGUAS	47
ANEXO 3: LISTA DE PAÍSES OU TERRITÓRIOS DE QUE A UIT NÃO FORNECE DADOS	48
ANEXO 4: RESULTADOS PARA TODAS AS LÍNGUAS	49

LISTA DE TABELAS E FIGURAS

TABELAS

Tabela 1 : Os 2 tipos de ponderação usados.....	6
Tabela 2: Sensibilidade dos números da Índia sobre a percentagem de pessoas conectadas à Internet .	9
Tabela 3: Fatores e fórmula da Wikipédia	12
Tabela 4: Ponderação dos indicadores de conteúdo.....	13
Tabela 5: Descrição dos indicadores.....	15
Tabela 6 : Indicadores para as 15 principais línguas em termos de poder	16
Tabela 7 : Línguas ordenadas por percentagem de pessoas conectadas	17
Tabela 8 : Línguas ordenadas por capacidade.....	18
Tabela 9 : Línguas ordenadas por gradiente.....	19
Tabela 10: Presença de línguas na Wikipédia.....	20
Tabela 11: Presença da Wikipédia ordenada por valores de fórmula	20
Tabela 12: Avaliação de viés por indicador	22
Tabela 13 : Macroindicadores para as 15 línguas principais após ponderar os indicadores	24
Tabela 14: Classificado pelo número de artigos da Wikipédia	24
Tabela 15: Artigos da Wikipédia classificados por fórmula	26
Tabela 16: Número de wikibooks	27
Tabela 17: Número de citações (WikiQuote).....	27
Tabela 18: Número de Wikisources.....	28
Tabela 19: Número da Wikiversidade.....	28
Tabela 20: Número de entradas do Wikcionário	29
Tabela 21: Número de Wikinews	29
Tabela 22: Número de artigos em Wikivoyages	30
Tabela 23: Comparações de diferentes medidas de tráfego	31
Tabela 24: Primeiro método de correção de vieses.....	32
Tabela 25: Correção de vieses 2º método	33
Tabela 26: Resultados da correção de vieses	36

FIGURAS

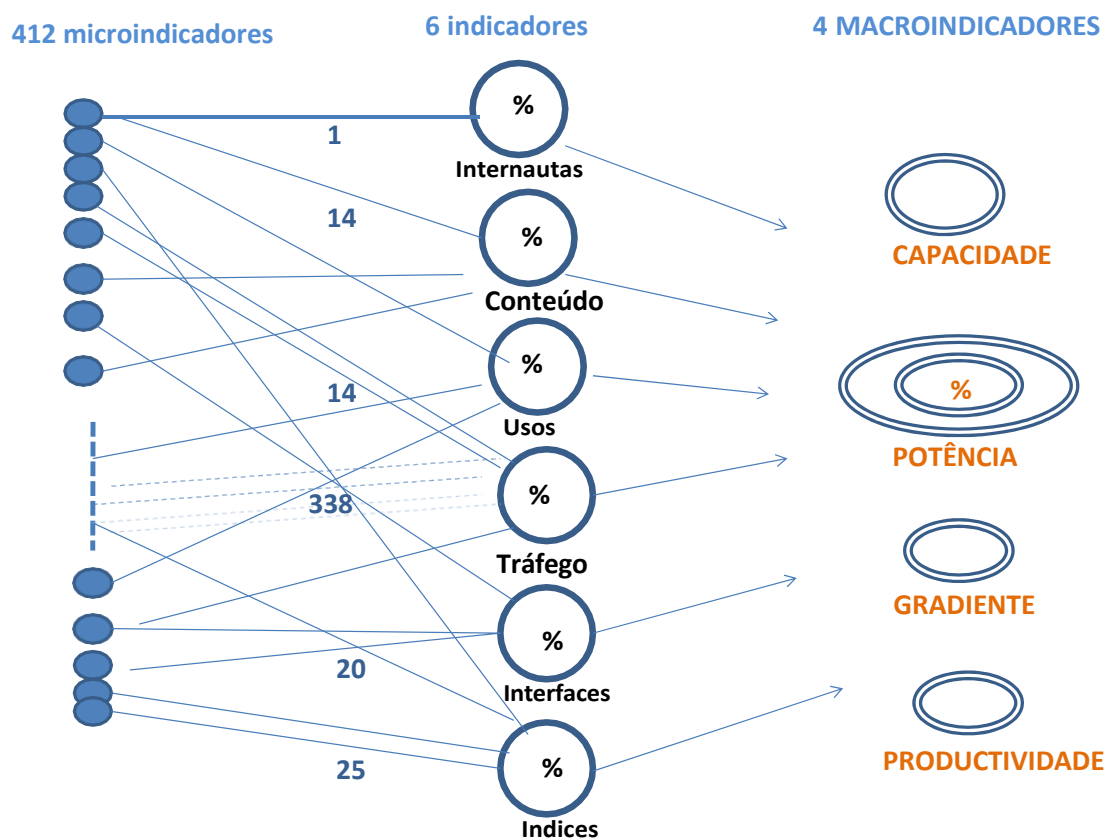
Figura 1: De microindicadores a macroindicadores.....	5
Figura 2: O processo de entrada / saída do modelo.....	6

CONTEXTO

A primeira edição deste método para produzir indicadores de presença de línguas na Internet foi realizada em 2017 e foi documentada sob o título “Uma abordagem alternativa para produzir indicadores de línguas na Internet” ([1]) e está acessível no site do Observatório em 4 versões linguísticas (inglês, francês, português e espanhol)¹. O leitor é convidado a consultar este documento antes de ler este artigo, que é complementar a esta primeira versão, e que apresenta o método e os resultados; este artigo apresenta as diferenças do método e os novos resultados.

Como um lembrete, o método se concentra nas 138 línguas cujo número de falantes L1² é maior que 5 milhões³ e produz indicadores para cada um deles, de acordo com o diagrama a seguir (cujos dados são atualizados para a segunda versão).

Figura 1: De microindicadores a macroindicadores



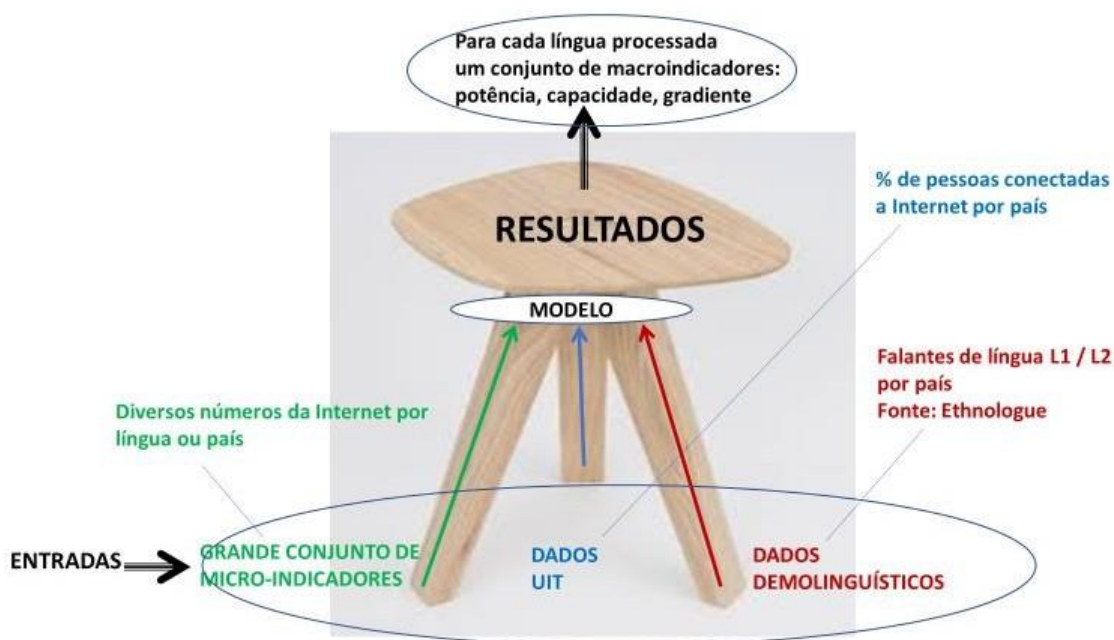
O método consiste em 3 tipos de entradas e dez sinalizadores de saída, conforme mostrado na figura a seguir.

¹ <http://funredes.org/lc2017>

² A convenção é chamar L1 de língua materna (ou primeira língua) e L2 de segunda língua, entendendo que é necessário um domínio suficiente de uma língua para ser qualificado dessa forma.

³ Na verdade, o total é 128: para permitir comparações com o estudo de 2017, há 10 línguas com menos de 5 milhões de falantes porque apareceram no estudo de 2017. Essas línguas são: awahadi, bielo-russo, bikol, bugis, dugri, armênio, kimbundo, lujia, flamengo ocidental e tailandês do sul.

Figura 2: O processo de entrada / saída do modelo.



O processo do modelo é baseado em mecanismos de ponderação capazes de transformar dados por país em dados por língua, técnicas de extrapolação para complementar fontes que oferecem valores para um número limitado de países e mecanismos de ponderação com a distribuição global de pessoas conectadas à Internet, a fim de produzir percentagens globais das várias fontes.

Tabela 1 : Os 2 tipos de ponderação usados.

	Demolinguística	Usuários da Internet por língua
ENTRADA	% por país ---> % por língua	% Critério ---> % mundial
SAÍDA	Dados por país	Dado em % de acordo com critérios específicos
PRODUÇÃO	Dados por língua	Dados populacionais L1 + L2
DADOS DE PONDERAÇÃO	Falantes L1 + L2 por país	% de pessoas conectadas à Internet por país
ESCOPO	Todas as fontes por país	Indicadores de índice e interface
PRESSUPOSTO IMPLÍCITO	Independência da língua no país	Taxa de modulação da conexão à Internet de acordo com critérios

O modelo é implementado em Excel em uma planilha de 7 MB com 17 guias correlacionadas, organizadas em torno dos 215 países considerados, as 138 línguas processados e os 412 microindicadores coletados. O modelo assim configurado permite verificar em uma fração de segundo o impacto de qualquer hipótese (incluindo análise prospectiva).

1. INTRODUÇÃO

Esta segunda versão do referenciado método de criação de indicadores de presença de línguas na Internet traz um conjunto de melhorias tangíveis que impactam positivamente na confiabilidade do método e reduzem vieses.

As principais melhorias decorrem da adoção do *Ethnologue Global Dataset 24*⁴, Março de 2021, que não só atualiza os dados demolinguísticos (o número de falantes de cada língua em cada país), mas também fornece os dados mais confiáveis sobre o assunto, mesmo que a precisão perfeita seja inacessível, e além disso, nesta última versão oferece a primeira fonte histórica para a distribuição do número de falantes de L2 por país, para cada língua.

2. DIFERENÇAS DA PRIMEIRA VERSÃO

Muitas diferenças foram feitas no método ou nas fontes em relação à versão 1, com o objetivo de melhorar a qualidade do método e dos produtos.

2.1 Adoção do conjunto de dados Ethnologue 24 como uma fonte demolinguística

A maior parte da fonte fornecida pelo Ethnologue está na forma de uma matriz Excel de 11.500 linhas no seguinte formato: “ISO639⁵, Nome da língua, nome do país, número de falantes L1, número de falantes L2, além de um grande número de parâmetros associados não usados para este método”.

Para obter o formato exigido pelo modelo (uma matriz com todos os países considerados na coluna e todas as línguas consideradas nas linhas), um conjunto de etapas cuidadosas foi implementado com o apoio de diferentes macros escritas para Excel. Uma das etapas mais complexas foi mesclar todos os dados de língua pertencentes à mesma macrolíngua. Este processo envolveu 60 macrolínguas agrupando 434 línguas diferentes⁶ (ver detalhes no anexo 2).

Após concluir esta etapa, o processo consistiu em reduzir a lista de línguas para manter apenas aqueles de que o modelo trata⁷, adicionando cuidadosamente todos os dados restantes por país em uma única linha "RESTO".

É importante compreender que a adoção de dados do Ethnologue implica na aceitação das regras de apresentação, que se baseiam em considerações puramente linguísticas:

- Agrupamento de macrolínguas⁸
- Lista de países e denominação correspondente em inglês.

A lista de países tratados pelo Ethnologue é maior do que a da UIT⁹ para o fornecimento de taxas de conexão à Internet por país: a UIT, como entidade das Nações Unidas, não separa, por exemplo, a

⁴ <https://www.ethnologue.com/product/ethnologue-global-dataset-0>

⁵ O código ISO de 3 caracteres atribuído a cada um das 7.486 línguas identificados.

⁶ Por exemplo, o macrolíngua árabe inclui 29 línguas, como o árabe egípcio ou o árabe marroquino.

⁷ Neste ponto, 138 línguas com mais de 5 milhões de falantes L1.

⁸ Um exemplo significativo é o caso da macrolíngua servo-croata, cuja definição agrupa, em ordem alfabética, bósnio, croata, montenegrino e sérvio. Esse reagrupamento forçado não atende aos critérios geopolíticos e pode até ser considerado controverso desse ponto de vista. Além disso, como algumas fontes separam claramente as línguas e os países envolvidos, existe o risco de erro nos resultados, mesmo que a entrada da fonte tenha sido transformada para levar esta situação em consideração (o risco surge quando os números não deveriam ser adicionados, mas na média, como no medidor de profundidade da Wikipédia).

⁹ A União Internacional de Telecomunicações (<http://itu.int>), a agência das Nações Unidas que fornece estatísticas sobre telecomunicações, incluindo a porcentagem de pessoas conectadas à Internet por país.

Martinica da França). Neste caso, a regra da UIT é a que devemos aceitar e a exigência tem sido compilar cuidadosamente os dados do Ethnologue para os 29 países não considerados pela UIT (para a lista completa, ver Anexo 3) em uma única coluna “Outros países”.

2.2 Gestão de L2 e multilinguismo

A inclusão dos dados mais recentes do Ethnologue no modelo permitiu, como subproduto, eliminar o grande viés do método ligado ao tratamento de L2. Pela primeira vez, existe uma fonte confiável que complementa, para cada língua, o número de falantes L1 por país com o número de falantes L2 por país. Na versão 2017, os valores de pessoas L2 conectadas foram calculados a partir do número total de falantes L2 no mundo aplicando a mesma taxa de conectividade obtida pelo modelo para falantes L1. Um viés significativo se deve ao fato de que, para certas línguas principais (como o francês e o inglês), uma alta proporção de falantes de L2 pertence a países em desenvolvimento, onde a taxa média de conexão é muito mais baixa do que a obtida em média para falantes de L1.

Outra consequência positiva da utilização dos dados do Ethnologue é a possibilidade de obter um “dado oficial” do multilinguismo. A razão mundial $(L1+L2)/L1$ foi estabelecida na edição de 2017 projetando os dados disponíveis para os países tratados: o resultado foi em torno de 1,25. Agora o número é indiretamente fornecido pelos dados do Ethnologue e seu valor é 1,43.

Os dados do Ethnologue são os seguintes:

- ✓ População mundial (total mundial de falantes L1): 7.231.699.136
- ✓ Falantes do Total Mundial L1+L2: 10.361.716.756
- ✓ A "taxa de multilinguismo" é, portanto, $10.361.716.756 / 7.231.699.136 = 1,4328$
(Em outras palavras, 43% da população mundial é pelo menos bilíngue).

Este número de 43% é muito melhor do que os 25% usados na primeira versão e não é um elemento anedótico do modelo, mas um dos elementos-chave. Como mostra o primeiro estudo, o viés mais comum e crítico no número de línguas oferecidos é que eles não consideram corretamente os falantes L2 (um problema que é totalmente expresso na Internet, onde a maioria dos usuários usam sua segunda língua e onde muitos sites são multilíngues¹⁰) Não prestar atenção a isso leva a enormes erros, muitas vezes ocultos "no resto das línguas", uma vez que as percentagens mundiais a serem calculadas para uma população de 10 bilhões (falantes L1 + L2) são calculadas em um total de 7 bilhões (a população mundial).

Nesta segunda edição do método, o princípio de medir tudo em termos da população L1+L2 (ao invés da população mundial) foi totalmente adotado para garantir a precisão dos resultados. Por esse motivo (e também por outras melhorias), a comparação entre os resultados de 2017 e os de 2021 deve ser feita com cautela. Na verdade, todos os macroindicadores, *potência*, mas também *capacidade* e *gradiente* agora seguem esta regra de serem calculados pela população L1+L2 em vez de pela população L1 e, portanto, aparecerão com valores mais baixos em comparação com a versão 2017.

¹⁰ De fato, os 6 indicadores abrangidos pelo estudo são multilíngues por natureza: os utilizadores da Internet visitam sites e geram tráfego nas diferentes línguas que falam, frequentemente os sites são multilíngues, as interfaces são multilíngues, os serviços de tradução abrangem diferentes línguas.

2.3 Fonte para pessoas conectadas à Internet

Até 2017, a ITU fornecia uma atualização anual de seus dados¹¹ sobre a porcentagem de pessoas que usam a Internet por país, incluindo suas próprias estimativas quando não havia fonte oficial para alguns países. Esses dados, que são um elemento central do método, foram considerados entre os mais confiáveis. Infelizmente, depois de 2017, a ITU decidiu parar de fornecer suas próprias estimativas, deixando muitos países (quase todos os países em desenvolvimento¹²) com os números antigos de 2017 a 2021.

Isso representou um sério problema para este estudo e, após algumas interações, levou à decisão de violar, neste caso, um princípio fundamental neste tipo de trabalho estatístico: nunca modificar os dados das fontes.

O Banco Mundial fornece seus próprios números¹³ para o mesmo indicador, que são claramente retirados da UIT, mas em vários casos excedem a limitação atual e propõem novos números onde a UIT deixou os dados de 2017. Isso é um progresso, mas muitos países ainda permanecem fora da atualização de 2017, que teria um impacto negativo nas línguas faladas nesses países e nos impede de perceber um progresso potencial.

Finalmente, foi decidido usar os dados do Banco Mundial para complementar os dados da UIT e, para os muitos casos desatualizados restantes, bem como realizar, para cada país em questão, uma pesquisa na Internet para dados confiáveis e fornecer estimativas baseadas, salvo indicação em contrário, na progressão linear aproximada de dados de anos anteriores.

Um caso continua a ser um problema: a Índia tem um número oficial de 20,1% em 2021, enquanto a estimativa da UIT em 2017 era de 32% e muitas fontes relatam um boom da Internet neste país com números em torno de 50%¹⁴! Não tendo obtido uma resposta da fonte oficial ou dos colegas indianos consultados, foi decidido, devido à importância primordial da Índia no contexto do estudo¹⁵, excepcionalmente violar um princípio ainda mais forte: o de não modificar as fontes oficiais. O pressuposto é que o valor fornecido pelo *Ministério indiano de estatística e implementação de programas* se refere apenas a conexões fixas e não inclui conexões de Internet móvel. Com base nessa premissa, o valor conservador de 40% foi estabelecido. Observe que a sensibilidade deste dado aos resultados não é marginal. Abaixo estão os diferentes resultados finais para hindi e bengali de acordo com o valor da porcentagem de pessoas conectadas à Internet na Índia.

Tabela 2: Sensibilidade dos números da Índia sobre a porcentagem de pessoas conectadas à Internet

% pessoas conectadas na Índia	20,08%	30%	40%	50%
Potência do hindi (alcance)	2,42% (10)	2,91% (8)	3,38% (5)	3,81% (4)
Potência do bengali (alcance)	0,75% (17)	0,82% (15)	0,88% (15)	0,95% (14)

¹¹ <https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2021/PercentIndividualsUsingInternet.xlsx>

¹² Apenas 80 países forneceram dados oficiais em 2019.

¹³ Fonte: <https://data.worldbank.org/indicator/IT.NET.USER.ZS>

¹⁴ Veja por exemplo <https://www.statista.com/statistics/255146/number-of-internet-users-in-india/> ou então https://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users

¹⁵ Com 34 línguas que fazem parte da lista de línguas consideradas, incluindo línguas importantes como o hindi e o bengali.

2.4 Gestão de fontes para microindicadores

Todo o processo de gestão das fontes dos microindicadores é a tarefa mais pesada e difícil do projeto, com elevado consumo de recursos humanos. Muitas etapas são necessárias:

1. Para cada indicador, pesquisar fontes na Internet.
2. Selecionar as fontes com base na confiabilidade e aplicabilidade do processo.
3. Coletar as fontes selecionadas em um formato que permita a integração automática no modelo.
4. Integrar as fontes coletadas no modelo e associar um tópico.
5. Avaliar o viés da fonte.

No Anexo 1, é apresentada a lista completa de fontes para cada indicador.

Para a execução da etapa 4, os dados devem ser transformados em formato Excel, com os nomes dos países e línguas correspondentes aos do modelo e na mesma ordem sequencial.

Para a etapa 3, todas as fontes são coletadas de um URL específico (consulte o Anexo 1 para obter a lista completa de URLs) e a maioria das fontes é fornecida no formato HTML. Algumas fontes estão no formato PDF e um subconjunto limitado (principalmente ITU e Banco Mundial) está no formato Excel, que visa transformar todas as fontes. O processo de conversão de PDF para o formato Excel pode ser relativamente simples na maioria dos casos, mas em alguns casos há uma incompatibilidade e alguns ajustes são necessários, como passar por um formato DOC intermediário primeiro.

O processo de transformação de HTML para Excel pode muitas vezes se transformar em um pesadelo real que requer muita imaginação e ajustes, incluindo, em alguns casos, a necessidade de ir buscar os dados dentro do código-fonte HTML e tentar a partir daí construir uma tabela usando a função de conversor do Excel, após limpar o código HTML.

Em um número crescente de casos, a fonte oferece acesso geográfico aos dados (mapas clicáveis) que, exceto quando o número de países ou línguas é limitado e a cópia manual não é muito pesada, torna o processamento automático impossível ou exige terceirização para uma pessoa de confiança para trabalho de coleta manual, que é tedioso e requer grande concentração e disciplina para evitar erros. Portanto, a coleta de dados de tráfego que incluiu centenas de microindicadores foi terceirizada.

O crédito deve ser dado a instituições (geralmente organizações internacionais ou ONGs) que fornecem os dados em um formato utilizável por computador (a Wikimédia fornece, por exemplo, em sua versão em inglês, tabelas HTML que são diretamente transformadas em formato Excel sem perda de estrutura).

Obter uma cópia da fonte no Excel ou em um formato compatível (geralmente uma matriz de nomes de países ou línguas com valores ou porcentagens associadas) não é o fim do processo. Com 250 países e centenas de línguas para processar e no lugar do código ISO inequívoco, o uso comum de nomes literais que podem estar em diferentes línguas e em grafias não padronizadas, a integração dos dados no modelo não pode ser feita à mão. Duas macros foram escritas para este processo, que em ambos os casos requerem uma configuração recursiva¹⁶ para integrar as diferentes grafias. As saídas dessas macros são arquivos Excel que podem ser usados diretamente para copiar as fontes em sua totalidade ou linha por linha para o arquivo Excel do modelo. Além da enorme economia de tempo, este método garante a obtenção dos dados sem erros.

Observe-se também que o manuseio de macrolíngua tornou esse processo ainda mais complexo, porque o agrupamento de língua deve ser feito nos dados de origem antes do processamento pela macro. Para citar alguns exemplos, as ocorrências frequentes de árabe egípcio ou marroquino nas fontes foram acumuladas na macrolíngua árabe e as de sérvio, bósnio, croata e montenegrino foram fundidas em servo-croata (o número de casos semelhantes é bastante elevado). Para o processamento manual das saídas de macro desconhecidas, a página Ethnologue que descreve cada código de língua foi usada como suporte.¹⁷

2.4.1 ÍNDICE

O prazo chegou muito cedo durante a produção da versão de 2017 e este indicador se mostrou insuficiente com uma única fonte fornecendo 5 microindicadores. Nesta ocasião, foi prestada a devida atenção e efetuada uma recolha de dados quase exaustiva para este indicador. Foi incluída uma grande variedade de parâmetros que caracterizam o progresso dos países na sociedade da informação, agora com 25 microindicadores, desde estabilidade elétrica até inteligência artificial, governança e muitos outros parâmetros (ver Anexo 1 para toda a gama).

2.4.2 CONTEÚDO

Conforme explicado acima, as fontes de dados linguísticos na Internet são extremamente escassas e, portanto, este indicador depende muito das estatísticas excepcionais da Wikimédia. No entanto, no capítulo sobre vieses do documento é estabelecido que a presença de línguas na Wikimédia não é proporcional à sua presença na Web e, portanto, não pode ser considerada um indicador confiável da distribuição linguística do conteúdo da Internet.

Uma forma de equilibrar os resultados das estatísticas da Wikimédia foi introduzida no modelo, mas o diagnóstico doloroso é que o conteúdo ainda é o indicador mais fraco deste método, embora seja um elemento muito sensível (as mudanças relativamente marginais nos valores deste indicador podem ter um grande impacto nos valores dos macroindicadores). Embora um dos principais objetivos do projeto seja, é claro, conhecer a distribuição linguística do conteúdo da Internet, há uma dificuldade frustrante em pesar corretamente o conteúdo da Internet e, neste caso, o macroindicador holístico pode permanecer a melhor aproximação¹⁸.

¹⁶ O processo recursivo termina quando o processo de origem não produz mais uma grafia desconhecida.

¹⁷ <https://www.ethnologue.com/language/srp>

¹⁸ Como mostra a primeira edição, o esforço louvável da W3Techs para obter números de conteúdo é caracterizado por vieses muito significativos e em muitos níveis (o mais forte, mas não o único, é a falta de consideração pelo

Para tentar controlar melhor a influência excessiva dos números da Wikimédia neste indicador, duas decisões foram tomadas. O primeiro refere-se exclusivamente à Wikipédia: em vez de ter um indicador para cada um dos tipos de dados fornecidos (número de artigos, editoriais ativos, edições e profundidade ¹⁹) uma fórmula foi colocada em prática para definir um único microindicador:

$$W(Li) = \text{Artigos}(i) \times \text{Edições}(i) \times \text{Editores}(i) \times \text{Profundidade}(i) / L1 + L2(i)^2$$

Esta fórmula expressa com mais precisão a atividade geral da Wikipédia por língua, reduzindo a proporção de línguas para os quais os bots, em vez de humanos, são usados para criar artigos copiando / traduzindo²⁰ e raramente os atualiza depois disso.

A tabela a seguir mostra como essa fórmula reflete melhor a realidade. A última coluna (presença), que é a relação entre o número de artigos e a população L1 + L2 (número de artigos por locutor), é uma demonstração clara de porque a presença de línguas na Wikipedia não é um bom indicador da presença global das línguas na Internet.

Observe que o valor da profundidade para o vietnamita não foi preenchido e colocamos o valor 1 para evitar uma fórmula nula.²¹

Tabela 3: Fatores e fórmula da Wikipédia

Língua	Artigos	Modificações	Usuários ativos	Profundidade	FÓRMULA	PRESENÇA
Inglês	6332139	1027716498	125399	1073	481775	0,47
Cebuano	5853095	32075254	186	2	275	36,71
Sueco	3050759	49330695	2148	12	22759	23,37
Alemão	2593827	212207089	18119	93	50897	1,92
Francês	2342875	183969129	18054	242	26424	0,88
Holandês	2060512	59302602	3933	17	13742	8,45
Russo	1736736	115035192	10425	137	4286	0,67
Italiano	1703284	121418801	8085	172	62435	2,51

multilinguismo e o fato de sites multilíngues que incluem o inglês provavelmente só sejam contados em inglês). Assim, esta fonte projeta valores para conteúdo em inglês extremamente exagerados (acima de 50% enquanto a realidade provavelmente está hoje abaixo de 25%). A falta de fontes múltiplas, portanto, mantém o mito na mídia de que mais da metade dos sites são em inglês. Este foi o caso entre 2007 e 2009 (ver [3]), mas desde o crescimento exponencial de chinês, hindi, árabe, turco, bengali, vietnamita, urdu, persa e marata, para citar os que estão nos primeiros 20 lugares e que juntos representam quase 28% do conteúdo, a situação mudou radicalmente e o inglês representa apenas um quarto do conteúdo. Entre 2000 e 2007, o mito persistente era que o inglês ocupava 80% da Web e essa desinformação finalmente desapareceu após 2009 com a publicação pela UNESCO de relatórios (ver [3] e [4]) que estabeleceram uma presença do inglês em torno de 50%. Como o inglês poderia ter permanecido estável em 50% do conteúdo por 14 anos com todas as mudanças que a Internet passou e um número de falantes de inglês conectados (L1 + L2) que agora está reduzido a 13% de todas as pessoas conectadas?

¹⁹ Citado da Wikimédia: *Profundidade*, definido como $[\text{Modificações} / \text{Artigos}] \times [\text{N}^\circ \text{ de artigos} / \text{Artigos}] \times [1 - \text{Stub-ratio}]$, é um indicador aproximado da qualidade de uma Wikipedia, que mostra a frequência com que seus artigos são atualizados. Não se refere à qualidade acadêmica.

²⁰ Sem essa fórmula, Cebuano, com um grande número de artigos, mas muito pouca profundidade, aparece com a maior pontuação de *gradiente*.

²¹ O baixo valor de profundidade reflete o fato de que 67% das postagens são criadas por bots, não humanos (fonte: https://www.wikiwand.com/en/Vietnamese_Wikipedia).

Espanhol	1698331	136390848	15694	210	2590	0,31
Polonês	1480982	63723938	4235	32	7742	3,64
Japonês	1277204	84188217	15173	85	8683	1.01
Vietnamita	1266628	65110373	2476	1	35	1,65
Chinês	1208732	66159632	8940	202	62	0,08
Árabe	1123561	54279052	5189	227	536	0,31
Ucraniano	1100281	32831286	2773	53	4823	3,32
Português	1067241	61371751	9508	176	1651	0,41

No capítulo que trata do viés, é apresentada uma análise aprofundada das estatísticas da Wikimédia.

A segunda decisão tomada para equilibrar a influência da Wikimédia no indicador de *conteúdo* é um sistema de ponderação implementado que dá mais importância ao T-Index do Translated²² que para toda a coleção de indicadores da Wikimédia. Jogar com diferentes configurações de fatores de ponderação tem mostrado a alta sensibilidade deste indicador, principalmente devido ao número muito baixo de fontes e ao fato de algumas línguas terem uma presença desproporcional em relação ao seu número de falantes.

A configuração de ponderação finalmente implementada é a seguinte:

Tabela 4: Ponderação dos indicadores de conteúdo

OBJETO	PESO
Amazon USA - Libras 2017 ²³	0,5
Fórmula da Wikipédia	1
Número de WikiBooks por língua	0,5
Artigos WikiQuote por língua	0,1
Número de artigos WikiSource por língua	0,1
Número de artigos da Wikiversidade por língua	0,1
Número de artigos do Wikcionário por língua	0,1
Número de artigos WikiNews por língua	0,1
Número de artigos do WikiVoyages por língua	0,1
Projeção do T-index para e-commerce em 2021	3

2.4.3 TRÁFEGO

Ponderou-se muito também sobre o trabalho para o indicador de tráfego, com muitas tentativas e erros. Em 2017, os dados do Alexa (porcentagem de tráfego por país para vários sites) foram considerados extremamente distorcidos em relação aos países asiáticos (especialmente Índia e China) e Brasil e um pouco distorcidos em favor do francês e do inglês. Quatro anos depois, a

²² Este índice, acessível em <https://translated.com/les-langues-qui-comptent>, é uma tentativa de medir o potencial das línguas no comércio eletrônico com base no número de usuários da Internet por língua, multiplicado pelo gasto estimado online. Ele usou números do Banco Mundial e da UIT e propôs uma projeção para 2021, que são os dados usados para o modelo. Além dos dados da Wikimédia, é uma das poucas fontes linguísticas sérias disponíveis na Internet.

²³ A falta de dados equivalentes acessíveis para 2021 e a situação com a Wikimédia levaram à decisão de manter este microindicador, embora não esteja atualizado.

coleta de dados de Alexa mostrou situações estranhas (algumas métricas não informam o tráfego no país onde o site foi criado²⁴) e a impressão de uma tendência a subestimar o tráfego dos países europeus e, por outro lado, a Índia parece bastante elevada em todos os lugares, não tanto na China.

Um estudo que comparou os dados de tráfego com os dados de assinatura de cinco grandes redes sociais confirmou as impressões intuitivas. Em resumo, o tráfego do Brasil parece muito subestimado em relação ao nível de assinatura, o mesmo para Alemanha, Espanha, França, Itália e Reino Unido; por outro lado, Índia, Japão e Coreia parecem estar com dados altamente exagerados. Consulte o capítulo sobre vieses para obter mais detalhes.

Diante desses resultados desanimadores, decidiu-se buscar uma ferramenta de medição alternativa. SimilarWeb parecia ser a melhor alternativa e os testes foram planejados antes de comprar uma assinatura. Infelizmente, não foi possível acessar aos dados de *tráfego* por país e, apesar das inúmeras tentativas de comunicação através de diferentes canais, incluindo o chat interativo da empresa, nunca se obteve a mais breve resposta.

Diante dessa situação de estagnação, outro provedor, o Semrush.com, foi julgado e foram coletados dados por país para os mesmos sites. O Semrush, ao contrário do Alexa, fornece, para cada site medido, resultados para todos os países, o que foi uma perspectiva interessante, eliminando a necessidade de extrapolação. Porém, acontece que em alguns casos o total é inferior a 100% (o que não é um problema) e outras vezes ultrapassa 100% (o que é um problema). Os números foram normalizados para serem 100% precisos usando uma regra de ponderação antes de entrar no modelo.

Depois de rodar o modelo, transformando os dados nacionais em dados linguísticos, os resultados não foram convincentes: o valor do chinês era muito baixo, o mesmo para o hindi e o árabe e para as “demais línguas”.

As diferenças extremas entre os resultados obtidos a partir dos dados do Alexa e do Semrush são um alerta sobre a confiabilidade dessas ferramentas e uma preocupação com os planos futuros de expandir o número de sites estudados e permitir a diferenciação temática para algumas línguas.

2.4.4 INTERFACES

A lista de línguas aceitas em interfaces de aplicativos importantes ou como destinos possíveis para serviços de tradução na web não apresenta nenhum problema particular. A lista de aplicativos selecionados pode ser vista no Anexo 1. Observe que, para reduzir a importância dos dados da Wikimédia no modelo, foi decidido remover as fontes da Wikimédia para este indicador.

²⁴ Por exemplo, theses.fr mostrou tráfego zero na França, assim como spip.net, um CMS usado principalmente na França.

2.4.5 USOS

Também não há dificuldades particulares para este indicador, além de encontrar dados gratuitos (principalmente número de assinantes por país) nas principais redes sociais. Por fim, foi possível incluir dados para os seguintes aplicativos: Facebook, Instagram, LinkedIn, Messenger, Pinterest, Reddit e Twitter. Além disso, algumas fontes diferentes das redes sociais foram incluídas, como o número de downloads do OpenOffice por país (veja a lista completa no Anexo 1).

2.5 Resumo dos indicadores

A tabela a seguir resume a descrição de cada um dos indicadores e como ele é construído a partir de microindicadores.

Tabela 5: Descrição dos indicadores

INDICADOR		TÉCNICO	CONFIABILIDADE / VIÉS
A: INTERNAUTAS	Indicador único baseado em dados da UIT e do Banco Mundial para % de pessoas conectadas por país, extrapolado quando faltam números.	Pesagem país -> língua sem extrapolação	Alta fiabilidade Viés muito marginal embora esteja aumentando devido à falta de atualização em muitos países.
B: USOS	Inclui 14 microindicadores com dados de 2021: - % telefone fixo + celular por país - % banda larga por país - Download cumulativo do OpenOffice - Facebook, Instagram, LinkedIn, Messenger, Netflix, Pinterest Twitter, YouTube, % de assinantes por país	Pesagem país -> língua ² extrapolado na proporção da velocidade de conectividade Média de micro-indicadores	Confiabilidade forte. Baixo viés.
C: TRÁFEGO	Alexa mediu o tráfego por país em uma seleção de 338 sites.	pesagem país -> língua extrapolado proporcionalmente Média truncada para 20%	Confiabilidade relativamente boa Mas as fortes tendências negativas de Alexa na Europa são confirmadas por comparações de tráfego e número de assinantes por país.
D: INDICE	Inclui 25 índices de várias fontes que medem parâmetros como: - E-governo - Acesso universal - E. participação - Infraestrutura geral (Ver Anexo 1 para a lista completa)	pesagem país -> língua extrapolado pelo método do quartil. Em seguida, a transformação para dados globais por ponderação percentual com ITU Média de microindicadores	Boa confiabilidade e viés marginal (dados subjetivos quantificados por um órgão competente).

E: CONTEÚDO	Inclui 13 microindicadores com ponderação associada. Translated T-Index, uma medida do potencial de comércio eletrônico de uma lista de línguas (2021) - Número de livros na Amazon (2017) - 11 microindicadores da língua Wikimédia: artigos, usuários ou editores; todos os indicadores da Wikimédia são sintetizados com uma fórmula.	Uso direto de dígitos ponderados por língua para equilibrar a importância da Wikimédia. Fusão de 4 indicadores da Wikipédia com uma fórmula. Média truncada para 20% do microindicador	Confiabilidade muito boa para Wikimédia e Amazon. Mas bastante enviesado devido à baixa presença de algumas das principais línguas asiáticas. O número de microindicadores deve ser aumentado para dar mais força à média.
F: INTERFACES (e línguas de tradução)	Inclui 23 microindicadores binários	% de presença nos 23 micro-indicadores. % ponderação mundial com números ITU.	Perfeito.

3. RESULTADOS

As tabelas a seguir apresentam os resultados mais elevados para as línguas em cada macroindicador, excluindo produtividade²⁵. A tabela abaixo mostra todos os resultados resumidos para as 15 línguas mais "poderosas" da Internet. Os resultados são expressos em porcentagens com base na população L1 + L2.

Con.M = Porcentagem Mundial de Pessoas Conectadas,

Pop.M = População Mundial

L. Con. = Porcentagem de falantes conectados.

Tabela 6 : Indicadores para as 15 principais línguas em termos de poder

	Con.M	Pop.M	TRÁFE.	L. Con.	USOS	CONT.	INTERF.	ÍNDICE	POTÊNC.	Capac.	Grad.
Inglês	15,30%	13,01%	37,44%	64,33%	27,92%	38,61%	21,73%	17,87%	26,48%	2.04	1,73
Chinês	17,65%	14,72%	7,79%	65,59%	5,47%	8,18%	25,07%	19,38%	13,92%	0,95	0,79
Espanhol	7,00%	5,24%	10,72%	73,08%	11,74%	5,42%	9,94%	7,59%	8,73%	1,67	1,25
Francês	3,00%	2,58%	2,64%	63,67%	3,75%	5,40%	4,26%	3,21%	3,71%	1,44	1,24
Hindi	4,26%	5,80%	4,81%	40,18%	3,16%	0,28%	4,03%	3,71%	3,38%	0,58	0,79
Português	3,05%	2,49%	1,42%	67,16%	5,53%	3,30%	3,85%	2,92%	3,35%	1,35	1,10
Russo	3,51%	2,49%	1,81%	77,20%	2,28%	3,38%	3,88%	3,78%	3,11%	1,25	0,88
Árabe	3,89%	3,53%	2,30%	60,14%	3,02%	2,05%	4,29%	3,01%	3,09%	0,88	0,80
Alemão	2,09%	1,30%	1,32%	87,65%	1,95%	5,84%	2,97%	2,98%	2,86%	2,19	1,37
Japonês	2,07%	1,22%	1,98%	92,62%	1,76%	3,55%	2,77%	3,01%	2,52%	2.07	1,22
Malaio	2,20%	2,36%	0,89%	51,00%	2,79%	0,79%	1,91%	1,99%	1,76%	0,75	0,80
Italiano	0,91%	0,66%	0,51%	75,65%	0,97%	3,39%	1,22%	1,20%	1,37%	2.09	1,51
Turco	1,21%	0,85%	1,03%	77,98%	1,59%	0,94%	1,43%	1,22%	1,24%	1,46	1.02
Coreano	0,93%	0,79%	0,93%	64,73%	0,99%	0,85%	1,10%	0,95%	0,96%	1,22	1.03
Bengali	1,14%	2,58%	1,22%	24,15%	1,13%	0,26%	0,72%	0,84%	0,88%	0,34	0,78
RESTO	31,79%	40,39%	23,19%		25,95%	17,77%	10,81%	26,34%	22,64%		
TOTAL	100%	100%	100%		100%	100%	100%	100%	100%		

²⁵Este indicador será reavaliado na seção Correção de Vieses. O indicador de *potência*, que integra todos os elementos, seria provavelmente, nesta fase, uma melhor aproximação à distribuição de conteúdo por língua, dados que até hoje são muito difíceis de obter com fiabilidade.

A linha RESTO representa o conjunto completo de todas as línguas do mundo, exceto as 15 línguas listadas na tabela. Deve ficar claro que a classificação em termos de potência privilegia as línguas que possuem o maior número de falantes. Os macroindicadores de capacidade e gradiente fornecem resultados independentes do número de falantes.

Memória:

Poder²⁶ foi definido como a média dos 5 indicadores.

Capacidade²⁷ é o valor da potência dividido pela % de falantes L1+L2

Gradiente²⁸ é o valor da potência dividido pela % dos falantes L1+L2 conectados

A tabela a seguir mostra as línguas mais conectadas.

Tabela 7 : Línguas ordenadas por porcentagem de pessoas conectadas

CLASSIFICAÇÃO POR INTERNAUTAS	Internautas	Capacidade	Gradiente
Dinamarquês	97,82%	2,19	1,22
Sueco	93,49%	2,61	1,53
Japonês	92,62%	2,07	1,22
Holandês	92,02%	2,26	1,34
Suíço-alemão	91,56%	1,21	0,72
Flamengo ocidental	90,43%	1,12	0,68
Finlandês	89,67%	3,42	2,09
Bávaro	87,68%	0,97	0,61
Alemão	87,65%	2,19	1,37
Hebraico	85,46%	5,24	3,35
Eslovaco	82,47%	1,30	0,86
Bielo-russo	82,27%	1,00	0,66
Tcheco	81,37%	1,70	1,14
Polonês	81,17%	1,88	1,26
Húngaro	79,92%	1,79	1,22
Tártaro	78,05%	0,87	0,61
Turco	77,98%	1,46	1,02
Servo-croata	77,78%	3,14	2,21
Grego	77,71%	1,75	1,23
Russo	77,20%	1,25	0,88
Cazaque	76,98%	0,90	0,64
Romeno	75,66%	1,18	0,86
Italiano	75,65%	2,09	1,51
Albanês	75,48%	1,12	0,81
Azerbaijani	74,76%	0,94	0,69
Napolitano-calabrês	74,39%	0,84	0,62
Espanhol	73,08%	1,67	1,25

²⁶A palavra potência foi usada em vez de ponderação para evitar confusão com o importante uso transversal da ponderação no método. Representa a presença absoluta de uma língua na Internet, integrando todos os fatores.

²⁷Capacidade é a presença relativa de uma língua na Internet, independentemente do número de falantes; indica o dinamismo de uma língua na Internet.

²⁸O gradiente indica o dinamismo dos falantes conectados; o termo gradiente foi escolhido, o que expressa uma derivada e, portanto, uma tendência, porque um forte gradiente é uma promessa de aumento de capacidade.

Curdo	73,02%	0,89	0,67
Búlgaro	70,34%	1,18	0,92
Armênio	69,86%	1,41	1,11
Vietnamita	69,04%	1.07	0,85
Guarani	68,83%	0,64	0,51
Português	67,16%	1,35	1,10

A tabela a seguir é ordenada por *capacidade*.

Tabela 8 : Línguas ordenadas por capacidade

CLASSIFICAÇÃO POR CAPACIDADE	Usuários de internet	Capacidade	Gradiente
Hebraico	85,46%	5,24	3,35
Finlandês	89,67%	3,42	2.09
Servo-croata	77,78%	3,14	2,21
Sueco	93,49%	2,61	1,53
Holandês	92,02%	2,26	1,34
Alemão	87,65%	2,19	1,37
Dinamarquês	97,82%	2,19	1,22
Italiano	75,65%	2.09	1,51
Japonês	92,62%	2.07	1,22
Inglês	64,33%	2.04	1,73
Polonês	81,17%	1,88	1,26
Húngaro	79,92%	1,79	1,22
Grego	77,71%	1,75	1,23
Tcheco	81,37%	1,70	1,14
Espanhol	73,08%	1,67	1,25
Turco	77,98%	1,46	1.02
Francês	63,67%	1,44	1,24
Armênio	69,86%	1,41	1,11
Português	67,16%	1,35	1,10
Eslovaco	82,47%	1,30	0,86
Russo	77,20%	1,25	0,88

E, por fim, a última tabela, ordenada por *gradiente*, destaca o dinamismo das pessoas conectadas. A presença na terceira posição do malgaxe²⁹ é uma consequência do dinamismo de seus falantes em alguns indicadores da Wikimédia.

²⁹Tal classificação para o malgaxe, uma língua com menos de 10% de falantes conectados e uma habilidade muito baixa, pode causar uma surpresa legítima: é o resultado de um "acidente matemático" devido a uma presença extremamente desproporcional na língua dos indicadores de conteúdo e é um sintoma da fraqueza desse indicador que é discutido a seguir.

Tabela 9 : Línguas ordenadas por gradiente

CLASSIFICAÇÃO DO GRADIENTE	Internautas	Capacidade	Gradiente
Hebraico	85,46%	5,24	3,35
<i>Servo-croata</i>	77,78%	3,14	2,21
<i>Malgaxe</i>	9,79%	0,40	2,21
Finlandês	89,67%	3,42	2,09
Inglês	64,33%	2,04	1,73
Sueco	93,49%	2,61	1,53
Italiano	75,65%	2,09	1,51
Alemão	87,65%	2,19	1,37
Holandês	92,02%	2,26	1,34
Polonês	81,17%	1,88	1,26
Espanhol	73,08%	1,67	1,25
Francês	63,67%	1,44	1,24
Grego	77,71%	1,75	1,23
Dinamarquês	97,82%	2,19	1,22
Húngaro	79,92%	1,79	1,22
Japonês	92,62%	2,07	1,22
Tcheco	81,37%	1,70	1,14
Armênio	69,86%	1,41	1,11
Português	67,16%	1,35	1,10

Além do previsível, que as línguas nacionais de países reconhecidos por suas políticas proativas em prol da sociedade da informação ocupem os primeiros lugares, é notável que várias línguas se classifiquem acima do inglês, apesar de sua vantagem estratégica na Internet (língua de escolha de conteúdos multilíngues e o fato de muitos acreditarem que é a língua franca da Internet).

Estes resultados devem ser tomados com algumas reservas devido aos vieses mencionados no documento, em particular as dificuldades com o indicador de conteúdo, cujas variações podem ter um impacto considerável nestes macroindicadores³⁰.

4. ANÁLISE DOS RESULTADOS

Embora as comparações com os resultados de 2017 devam ser feitas com cautela devido ao tamanho e natureza das mudanças (em particular, a opção de expressar percentuais em relação ao total da população mundial L1+L2), alguns fenômenos podem ser destacados.

O esperado crescimento do hindi que em breve poderá competir com o francês pelo 4º lugar e a aparição do turco na lista das línguas mais poderosas. Como esperado também, as diferenças entre as línguas após o francês e o hindi são muito pequenas para considerar que os resultados estão além do intervalo de confiança: português, russo, árabe e alemão. No entanto, a demografia pode, em um futuro próximo, separar as respectivas posições no ritmo do estreitamento da brecha digital.

³⁰ Antes da introdução da fórmula da Wikipédia e da ponderação da Wikimédia, o cebuano, a segunda língua em termos de número de artigos da Wikipédia, próximo ao inglês, com um número de artigos duas ordens de magnitude maior que seu número de falantes, aparecia na primeira linha.

Quanto aos macroindicadores independentes do número de falantes, o aparecimento do servo-croata deve ser considerado com cautela devido ao risco de erros na gestão das fontes decorrentes da decisão de adoção da classificação do Ethnologue para macro-línguas. E, claramente, o indicador de conteúdo, e sua forte dependência das estatísticas da Wikimédia, apesar do esforço feito para combatê-lo, favorece claramente as línguas cujos falantes investiram na presença da Wikimédia. Veja a tabela abaixo para essas línguas, primeiros ordenados pela razão 1000 x Número de artigos / falantes de L1 + L2 e depois ordenados pelo resultado da fórmula implementada (fator).

Tabela 10: Presença de línguas na Wikipédia

Língua	Artigos	Edições	Usuários ativos	Profundidade	FATOR	% FATOR / L1 + L2	% FATOR / CON	ART / L1 + L2
Sueco	3050759	49330695	2148	12	22759	1,74	1,86	233,68
Finlandês	512026	19813368	1752	40	21354	3,70	4,13	88,74
Holandês	2060512	59302602	3933	17	13742	0,56	0,61	84,51
Servo-croata	1514114	78699318	1959	92	53779	2,69	3,46	75,77
Bielo-russo	281379	6093511	384	61	2620	0,67	0,81	71,87
Dinamarquês	267641	10777444	767	64	4486	0,80	0,82	47,64
Húngaro	489514	23958462	1561	59	6871	0,55	0,69	39,04
Polonês	1480982	63723938	4235	32	7742	0,19	0,23	36,44
Tcheco	484445	20095461	2242	46	5593	0,42	0,51	36,16
Ucraniano	1100281	32831286	2773	53	4823	0,15	0,23	33,16
Búlgaro	273163	11023721	789	27	942	0,11	0,16	33,10
Hebraico	298053	31660591	3335	258	92147	9,82	11,49	31,75
Italiano	1703284	121418801	8085	172	62435	0,92	1,22	25,10
Alemão	2593827	212207089	18119	93	50897	0,38	0,43	19,21
Japonês	1277204	84188217	15173	85	8683	0,07	0,07	10,11
Persa	816984	32472834	5416	172	3534	0,04	0,07	9,77
Francês	2342875	183969129	18054	242	26424	0,10	0,16	8,78
Inglês	6332139	1027716498	125399	1073	481775	0,36	0,56	4,70

A tabela abaixo mostra claramente porque algumas línguas, como hebraico, finlandês e servo-croata, obtiveram vantagem nos resultados finais.

Tabela 11: Presença da Wikipédia ordenada por valores de fórmula

Língua	FATOR	% FATOR / L1 + L2	% FATOR / COM
Hebraico	92147	9,82	11,49
Finlandês	21354	3,70	4,13
Servo-croata	53779	2,69	3,46
Sueco	22759	1,74	1,86
Italiano	62435	0,92	1,22
Dinamarquês	4486	0,80	0,82
Bielo-russo	2620	0,67	0,81
Húngaro	6871	0,55	0,69
Holandês	13742	0,56	0,61
Inglês	481775	0,36	0,56
Tcheco	5593	0,42	0,51
Alemão	50897	0,38	0,43
Polonês	7742	0,19	0,23

Ucraniano	4823	0,15	0,23
Búlgaro	942	0,11	0,16
Francês	26424	0,10	0,16
Japonês	8683	0,07	0,07
Persa	3534	0,04	0,07

Essas considerações conduzem naturalmente à discussão de vieses.

5. ANÁLISE DE VIESES

Existem três categorias principais de vieses que podem afetar os resultados:

- Vieses específicos do método
- Vieses de seleção de fonte
- Vieses de fontes

5.1 Vieses específicos do método

Um dos principais vieses do método, que consiste em atribuir o mesmo valor percentual de falantes L1 conectados à Internet para falantes L2, foi eliminado com a mudança para os dados Ethnologue, graças à distribuição dos falantes L2 por país. Esse viés significativo afetou particularmente as línguas com uma grande população de L2 em países com baixa taxa de conectividade (francês e inglês). Este é um avanço fundamental para a confiança nos números produzidos pelo modelo estabelecido.

O segundo viés do método é considerar que, em um determinado país, todos os falantes têm a mesma porcentagem de conectividade (ou seja, a porcentagem nacional de pessoas conectadas à Internet se aplica da mesma forma a todos os falantes quando é claramente uma média). Esse viés proíbe a distinção entre falantes de diferentes línguas dentro de um país com o método (por exemplo, falantes de catalão na Espanha recebem a mesma porcentagem de conectividade que falantes de espanhol e nenhuma vantagem de diferenciação pode ser analisada (é o mesmo com o crioulo de Martinica na França ou com as várias línguas da Índia). É intuitivamente entendido que esta hipótese não é verdadeira em muitos casos (a brecha digital nacional está frequentemente ligada a considerações linguísticas) e que o impacto deste viés é tanto mais forte quanto menor for a população considerada. Um efeito marginal é esperado se o método for aplicado a uma população de falantes de mais de 5 milhões (embora no caso da Índia isso possa não ser tão óbvio). O próximo lançamento do modelo, com previsão de término antes do final de 2021, tentará ampliar o limite para línguas com mais de um milhão de falantes.

Outros vieses marginais no modelo podem resultar da adoção de estruturas implícitas nas fontes principais. Por exemplo, a divisão em países foi derivada da classificação da UIT e não distingue certos territórios e, portanto, atribui a eles a mesma porcentagem do país de origem (se a taxa de conectividade de um território anexado for de fato muito menor do que a do país de origem, anexo que poderia ser o caso de Mayotte, as línguas específicas deste território, por exemplo, o kibushi, beneficiar-se-ão de um viés favorável).

5.2 Vieses de seleção de fontes

É evidente que existe um "viés de seleção", que não é específico da metodologia, mas pertence à aplicação do método, onde a decisão sobre a seleção das fontes favorece implicitamente critérios específicos da origem cultural do autor que inconscientemente ignora países muito longe de sua experiência. Isso pode se aplicar a cada um dos indicadores e ter um impacto específico no indicador de *tráfego*, onde a seleção do site tem grande influência, mesmo que o número de sites esteja na casa das centenas. O uso da média truncada de 20% foi implementado para reduzir esses vieses, após verificar que 20% era um grande intervalo capaz de eliminar a grande maioria dos resultados voltados para sites com forte localidade linguística.

5.3 Vieses das fontes

Os vieses decorrentes das fontes são analisados na tabela abaixo, classificando cada indicador com um valor de 0 (totalmente enviesado) a 20 (sem nenhum viés).

Tabela 10: Avaliação de viés por indicador

INDICADOR	AValiação	COMENTÁRIOS
INTERNAUTAS	19→16	Este indicador é derivado de um único microindicador. A principal fonte é a UIT. Em 2017, foi a fonte mais bem avaliada com 19/20, mas nesta versão a pontuação cai para 16 porque a UIT deixou de fornecer sua própria estimativa quando o país não produz dados oficiais. Os números da UIT foram complementados com os números do Banco Mundial e uma projeção linear de dados de anos anteriores foi estabelecida para os outros casos. Este indicador é fundamental no método porque é utilizado para ponderar os resultados em várias situações, porém, a análise fatorial mostrou que o impacto de uma pequena variação é moderado. Por exemplo, se a taxa de pessoas conectadas para o Brasil fosse definida para 80% em vez do valor real de 74%,
ÍNDICE	15→18	Este indicador é derivado de uma combinação de 25 microindicadores que avaliam diferentes parâmetros nacionais que caracterizam a sociedade da informação. As fontes são organizações internacionais, ONGs ou universidades. Vieses, se existirem, são marginais. O viés de seleção aqui é extremamente baixo porque estamos próximos da integridade de todos os microindicadores.
CONTEÚDO	5→8	Existem apenas 13 microindicadores para construir este indicador e 11 deles vêm da Wikimédia. A distribuição de conteúdo na web por língua é um continente oculto da Internet e as fontes existentes são extremamente raras e, muitas vezes, distorcidas. Infelizmente, o modelo não escapa dessa situação em seu estado atual. Como é amplamente baseado nas excelentes estatísticas da Wikimédia, o indicador tem o viés da Wikimédia, onde a presença de línguas asiáticas é muito menor do que sua

		<p>proporção na Internet. Obviamente, o viés de seleção neste caso, que depende muito das estatísticas da Wikimédia, é extremamente importante. Um sistema de ponderação foi posto em prática para reduzir esta dependência tanto quanto possível (o que em qualquer caso certamente não é suficiente, razão pela qual a classificação passou de 5 ‘muito baixo’ para 8 ‘insuficiente’). O viés do indicador de conteúdo é bastante sensível (ou seja, as variações têm forte impacto nos resultados), como mostra o experimento realizado simulando com o sistema de ponderação. Algumas ideias para tentar remediar esse problema serão implementadas na próxima edição. Enquanto isso, os preconceitos são superados "manualmente" usando certas técnicas (consulte Correção de vieses).</p>
TRÁFEGO	13→11	<p>Este indicador é derivado da medição do tráfego por país usando Alexa.com em uma seleção de 338 sites. Em 2017, a análise de viés mostrou que essa fonte estava fortemente enviesada em relação aos países asiáticos e ao Brasil. Em 2021, parece que o viés contra os países asiáticos foi corrigido (talvez até demais no caso da Índia!), mas novos vieses são detectados que agora afetam os países europeus. O viés de seleção é evidente neste caso e a próxima versão aumentará seriamente o número de sites medidos. A possibilidade de mesclar resultados de Semrush e Alexa em proporções iguais para conter vieses existentes deve ser explorada.</p>
INTERFACES	19	<p>Estes são dados objetivos (se uma língua está presente na interface de um aplicativo ou como um destino para um serviço de tradução online). Pode haver viés de seleção e a lista pode precisar ser expandida, mas seu impacto é marginal. Intuitivamente, podemos ver um aumento, em relação a 2017, no número de línguas suportadas em interfaces ou tradução; no entanto, continua sendo um "indicador radical" que omite a grande maioria das línguas do mundo e se concentra em um subconjunto muito limitado.</p>
USOS	12	<p>Este indicador é baseado principalmente em dados de assinatura de mídia social por país. Embora os dados coletados possam ser considerados confiáveis, o método implica um viés contra países não ocidentais com aplicativos alternativos ao Facebook, Twitter, LinkedIn, etc. A próxima campanha de medição tentará identificar populações alternativas de assinantes de aplicativos para equilibrar os resultados e tentar reduzir o viés. Durante este tempo, a correção de viés deve ser feita manualmente. O viés de seleção não existe realmente porque a seleção é ditada pela estreiteza das opções existentes. A próxima versão beneficiará de um pequeno orçamento para a base de dados comercial que deverá permitir expandir o número de microindicadores.</p>

Se a ponderação apresentada na tabela anterior for aplicada aos resultados na construção da média ponderada dos macroindicadores de *potência* (ao invés da média simples), de forma que a confiança relativa nos diferentes indicadores do modelo seja levada em consideração, os seguintes ajustes são anotados nos resultados, para serem comparados com os resultados anteriores, para levá-los em consideração ao corrigir os vieses.

Tabela11 : Macroindicadores para as 15 línguas principais após ponderar os indicadores

	POTÊNCIA.	Capac.	Grad.	POTÊNCIA.	Capac.	Grad.	Efeito
Inglês	24,23%	1,86	1,58	26,48%	2,04	1,73	---
Chinês	15,77%	1,07	0,89	13,92%	0,95	0,79	+++
Espanhol	8,80%	1,68	1,26	8,73%	1,67	1,25	+
Hindi	3,63%	0,63	0,85	3,38%	0,58	0,79	+++
Francês	3,62%	1,40	1,21	3,71%	1,44	1,24	-
Português	3,37%	1,36	1,10	3,35%	1,35	1,10	+
Árabe	3,28%	0,93	0,85	3,09%	0,88	0,80	++
Russo	3,24%	1,30	0,92	3,11%	1,25	0,88	++
Alemão	2,72%	2,08	1,30	2,86%	2,19	1,37	-
Japonês	2,51%	2,06	1,22	2,52%	2,07	1,22	
Malaio	1,87%	0,79	0,85	1,76%	0,75	0,80	++
Turco	1,27%	1,49	1,05	1,24%	1,46	1,02	+
Italiano	1,23%	1,88	1,36	1,37%	2,09	1,51	-
Coreano	0,97%	1,24	1,04	0,96%	1,22	1,03	
Bengali	0,91%	0,35	0,79	0,88%	0,34	0,78	+

5.3.1 Vieses da Wikipédia

As estatísticas da Wikipédia são impecáveis, no entanto, deve-se entender que embora seja uma das aplicações mais globais da Internet, ela mostra números para algumas línguas asiáticas que estão bem abaixo de sua presença relativa na web. A tabela a seguir compara as proporções entre o número de artigos da Wikipédia e o número de usuários da Internet. Grandes diferenças aparecem com valores anormalmente baixos para línguas asiáticas (com algumas exceções notáveis).

Tabela12: Classificado pelo número de artigos da Wikipédia

Língua	Artigos	% TOTAL ART.	Pesagem %	Art./L1+L2
Inglês	6332139	12,92%	0,28%	7
Cebuano	5853095	11,94%	22,16%	851
Sueco	3050759	6,22%	14,11%	250
Alemão	2593827	5,29%	1,16%	22
Árabe	2433772	4,97%	0,40%	11
Francês	2342875	4,78%	0,53%	14
Holandês	2060512	4,20%	5,10%	92
Chinês	1752600	3,58%	0,07%	2
Russo	1736736	3,54%	0,41%	9
Italiano	1703284	3,47%	1,51%	33
Espanhol	1698331	3,46%	0,19%	4
Servo-croata	1514114	3,09%	4,57%	97
Polonês	1480982	3,02%	2,20%	45

Japonês	1277204	2,61%	0,61%	11
Vietnamita	1266628	2,58%	1,00%	24
Ucraniano	1100281	2,24%	2,00%	52
Português	1067241	2,18%	0,25%	6
Malaio	936876	1,91%	0,23%	8
Persa	816984	1,67%	0,59%	15
Coreano	543656	1,11%	0,40%	10
Finlandês	512026	1,04%	5,36%	99
Húngaro	489514	1,00%	2,36%	49
Tcheco	484445	0,99%	2,18%	44
Romena	421153	0,86%	1,06%	2,3
Armênio	420677	0,86%	6,60%	156
Azerbaijani	420677	0,86%	1,06%	24
Turco	410954	0,84%	0,28%	6
Tártaro	299494	0,61%	3,42%	73
Hebraico	298053	0,61%	1,92%	37
Bielo-russo	281379	0,57%	4,34%	87
Búlgaro	273163	0,56%	2,00%	47
Dinamarquês	267641	0,55%	2,88%	49
Eslovaco	237210	0,48%	1,98%	40
Cazaque	228493	0,47%	1,05%	2,3
Grego	195481	0,40%	0,89%	19
Urdu	164062	0,33%	0,04%	3
Hindi	148545	0,30%	0,01%	1
Uzbeque	140894	0,29%	0,25%	9
Tâmil	138490	0,28%	0,10%	4
Tailandês	137351	0,28%	0,14%	3
Bengali	109438	0,22%	0,02%	2

Observe a presença de cebuano na segunda posição e a relativa presença do chinês e das línguas da Índia. É útil verificar uma porcentagem ponderada de acordo com o número de falantes L1+L2: o inglês não parece desproporcional e algumas línguas parecem ter uma presença forte em relação à sua população L1+L2, em ordem de importância: cebuano, sueco, armênio, finlandês, holandês, servo-croata, bielorrusso e tártaro.

A Wikimédia é o espaço virtual com maior diversidade linguística e o único que fornece estatísticas linguísticas claras e confiáveis sobre todas as suas atividades. Somando-se à centralidade de sua função na Web, é sem dúvida um indicador essencial quando se trata de conteúdo. Infelizmente, uma análise séria mostra que este espaço tão especial não poderia de forma alguma refletir uma indicação verdadeira da distribuição de conteúdo por língua na web. A importância das línguas na Wikimédia nem sempre está ligada à sua real importância no ciberespaço e algumas línguas investiram pesadamente neste espaço, independentemente de sua presença global na web.

Conforme explicado acima, o número de artigos não é um indicador confiável porque, para algumas línguas, foram implementados bots que criaram artigos a partir de traduções, que posteriormente permanecem sem atualização. Para controlar esse fenômeno, devemos atentar para o número de editores ativos, o número de edições e a profundidade, indicador criado para refletir o grau de atualização dos artigos. Uma fórmula para integrar esses fatores foi desenvolvida e apresentada acima. Os resultados ordenados por esta fórmula e apresentados como porcentagens são os seguintes:

Tabela 15: Artigos da Wikipédia classificados por fórmula

Inglês	53,96%
Hebraico	10,32%
Italiano	6,99%
Servo-croata	6,02%
Alemão	5,70%
Francês	2,96%
Sueco	2,55%
Finlandês	2,39%
Holandês	1,54%
Japonês	0,97%
Polonês	0,87%
Armênio	0,84%
Húngaro	0,77%
Tcheco	0,63%
Ucraniano	0,54%
Dinamarquês	0,50%
Russo	0,48%
Persa	0,40%
Bielo-russo	0,29%
Espanhol	0,29%
Português	0,18%
Árabe	0,16%
Romeno	0,13%
Búlgaro	0,11%
Coreano	0,10%
Turco	0,10%
Grego	0,07%
Eslovaco	0,04%
Cebuano	0,03%
Azerbaijani	0,02%
Malaio	0,02%
Tailandês	0,01%
Chinês	0,01%
Malaiala	0,00%
Cazaque	0,00%
Afrikaans	0,00%
Tártaro	0,00%

Bengali	0,00%
Mongol	0,00%
Tagalo	0,00%

Esta é claramente uma representação mais justa da realidade com a Wikipédia, prestando atenção equilibrada ao número de editores, edições e profundidades, e ponderada pelo número de falantes L1+L2. Tenha-se em mente que cebuano é penalizado com esta fórmula por sua política de uso de bots, mas que outra língua das Filipinas consegue ingressar na tabela: o tagalo. O domínio do inglês na Wikimédia também aparece mais claramente com essa abordagem.

A Wikimédia não se limita à Wikipédia e também existem estatísticas para cada um dos outros itens: WikiBooks, WikiQuote, WikiSource, Wikiversidade, Wiktionary, WikiNews e WikiVoyages, para os quais o número de artigos por língua está disponível. Para estes itens da Wikimédia, as fontes são apresentadas em termos absolutos, sem ponderação de acordo com o número de falantes, mostrando apenas os primeiros.

Tabela 1613: Número de wikibooks

Inglês	3851195	35,72%
Alemão	961696	8,92%
Francês	657991	6,10%
Português	473196	4,39%
Italiano	411671	3,82%
Polonês	403336	3,74%
Húngaro	401256	3,72%
Espanhol	396546	3,68%
Holandês	349987	3,25%
Vietnamita	256386	2,38%
Russo	205469	1,91%
Japonês	178783	1,66%
Árabe	174452	1,62%
Hebraico	164355	1,52%
Chinês	141302	1,31%
Finlandês	131314	1,22%
Persa	112964	1,05%
Malaio	89019	0,83%
Hindi	73969	0,69%

Tabela 14: Número de citações (WikiQuote)

Inglês	33897	14,28%
Italiano	30799	12,98%
Polonês	28960	12,20%
Russo	13148	5,54%
Tcheco	9263	3,90%
Persa	8495	3,58%
Alemão	7879	3,32%

Português	7443	3,14%
Espanhol	7116	3,00%
Servo-croata	7022	2,96%
Francês	5923	2,50%
Ucraniano	5798	2,44%
Eslovaco	4547	1,92%
Turco	4503	1,90%
Búlgaro	4389	1,85%
Hebraico	4202	1,77%

Tabela 15: Número de Wikisources

Francês	2609546	25,3%
Inglês	2204231	21,3%
Chinês	778716	7,5%
Bengali	722295	7,0%
Polonês	669381	6,5%
Russo	642705	6,2%
Alemão	431714	4,2%
Italiano	415032	4,0%
Tâmil	411502	4,0%
Hebraico	214947	2,1%
Sueco	84882	0,8%
Árabe	80708	0,8%
Multilíngue	78809	0,8%
Armênio	75487	0,7%
Português	73139	0,7%

Tabela 16: Número da Wikiversidade

Alemão	49011	36,9%
Inglês	38612	29,0%
Francês	17553	13,2%
Russo	5883	4,4%
Tcheco	5195	3,9%
Português	4692	3,5%
Italiano	4472	3,4%
Espanhol	2662	2,0%
Finlandês	1914	1,4%
Esloveno	1252	0,9%
Sueco	858	0,6%
Grego	644	0,5%
Japonês	207	0,2%

Tabela 20: Número de entradas do Wikcionário

Inglês	5923218	19,2%
Malgaxe	5466228	17,7%
Francês	3392407	11,0%
Chinês	1239843	4,0%
Servo-croata	1177979	3,8%
Russo	1002462	3,2%
Espanhol	885649	2,9%
Alemão	737337	2,4%
Holandês	686499	2,2%
Sueco	674872	2,2%
Polonês	649612	2,1%
Curdo	635201	2,1%
Lituano	616313	2,0%
Grego	462897	1,5%
Italiano	434058	1,4%
Coreano	398737	1,3%
Finlandês	374056	1,2%

É importante tentar entender o que aconteceu ao malgaxe e perguntar se sua classificação anormal na terceira posição no macroindicador gradiente invalida o método. Esta língua ocupa o segundo lugar no Wikcionário e mostra uma porcentagem do total de entradas de 17%, extremamente desproporcional em comparação com sua população (18 milhões de falantes) e muito mais ainda em comparação com seu número muito baixo de falantes conectados (1,8 milhões). Apesar do peso deste microindicador ter sido definido em 0,1 (o mesmo que toda a Wikimédia, exceto a fórmula da Wikipédia e Wikibooks), a desproporção é tão grande que afeta uma média ponderada com apenas 9 itens e, em cascata, o macroindicadores de *potência* e *gradiente*.

Tabela 21: Número de Wikinews

Inglês	21687	14,9%
Francês	20761	14,3%
Russo	17649	12,1%
Polonês	14357	9,9%
Espanhol	11312	7,8%
Chinês	8559	5,9%
Árabe	7578	5,2%
Servo-croata	5650	3,9%
Tcheco	5608	3,9%
Catalão	4056	2,8%
Tâmil	3363	2,3%
Sueco	3317	2,3%
Grego	3084	2,1%
Ucraniano	1738	1,2%
Romeno	1697	1,2%
Persa	1645	1,1%
Búlgaro	1562	1,1%

Português	1474	1,0%
Alemão	1386	1,0%

Tabela 17: Número de artigos em Wikivoyages

Inglês	28852	28,1%
Alemão	16545	16,1%
Persa	8674	8,5%
Italiano	7619	7,4%
Francês	7407	7,2%
Polonês	6946	6,8%
Russo	5438	5,3%
Holandês	3671	3,6%
Português	3624	3,5%
Chinês	2972	2,9%
Espanhol	2524	2,5%
Hebraico	2072	2,0%
Vietnamita	1624	1,6%
Sueco	1522	1,5%
Grego	1408	1,4%
Romena	917	0,9%
Ucraniano	779	0,8%

A diversidade dos resultados de acordo com cada tópico impossibilita uma conclusão sistemática da análise desses números, no entanto, algumas afirmações gerais podem ser feitas:

- O inglês está geralmente, mas nem sempre, em primeiro lugar, embora a proporção do inglês seja menos predominante do que na Wikipédia e permaneça entre 14% e 36%, com uma média de 23,5% (em comparação com 29,4% dos indicadores da Wikipédia)³¹.
- Francês e alemão pontuam alto na maioria dos indicadores da Wikimedia.
- Chinês, hindi, bengali e persa se destacam em algumas das rubricas.
- Certas línguas inesperadas aparecem no topo da lista para certos indicadores: malgaxe e tâmil (além de cebuano).

Em conclusão, a Wikimedia continua sendo de longe o lugar com maior diversidade linguística na Internet, com algumas línguas minoritárias com pontuação alta, mas dificilmente refletindo a verdadeira diversidade de conteúdo na Web. O inglês é predominante, mas não tanto quanto costumava ser. Em todos os casos, o método deve, prioritariamente, melhorar a qualidade do indicador de *conteúdo*, o que é preocupante no seu estado atual. Uma abordagem a explorar é identificar aplicativos semelhantes à Wikimedia que conseguiram ocupar um nicho em espaços linguísticos e de alguma forma introduzi-los nas estatísticas.

³¹ Esses percentuais referem-se ao número de artigos em inglês dividido pelo número total.

5.3.2 Vieses da Alexa

A tabela a seguir mostra os diferentes testes e comparações feitos entre Alexa e Semrush e, para Alexa, entre os dois anos de uso (2017 e 2021). Para Alexa 2017, os números de tráfego acima foram inseridos no modelo de 2021 para uma comparação justa, não influenciados por mudanças no modelo. A comparação não é feita com base nas entradas (por país), mas com base nos resultados do modelo (por língua); ou seja, a comparação é feita com o produto do modelo, inserindo como entrada cada um dos respectivos dados de tráfego. As comparações destacam (em vermelho na tabela) muitas anomalias.

Tabela 23: Comparações de diferentes medidas de tráfego

	SEMRUSH 2021	ALEXA 2021	2021 (SA) / S	ALEXA 2017	A21-A17 / A21
Inglês	52,50%	35,83%	32%	45,40%	-27%
Chinês	1,88%	7,67%	-308%	4,94%	36%
Espanhol	14,45%	10,14%	30%	7,53%	26%
Francês	4,48%	2,56%	43%	6,35%	-148%
Russo	1,88%	1,83%	3%	1,68%	8%
Alemão	2,61%	1,33%	49%	2,94%	-122%
Português	2,18%	1,46%	33%	1,63%	-12%
Árabe	1,02%	2,51%	-145%	2,54%	-1%
Hindi	1,26%	5,37%	-327%	1,60%	70%
Japonês	0,65%	1,94%	-198%	1,90%	2%
Malaio	0,68%	0,98%	-44%	1,23%	-27%
Italiano	0,89%	0,53%	41%	0,91%	-72%
Turco	0,60%	1,03%	-74%		
Polonês	0,47%	0,31%	3. 4%	0,63%	-100%
Coreano	0,50%	0,90%	-78%	0,72%	vinte%
RESTO	13,95%	25,34%	-82%	18,99%	25%
TOTAL	100,00%	100,00%	0%	100,00%	0%

- 1) É claro que Semrush e Alexa não refletem a mesma distribuição de tráfego por país para o mesmo conjunto de sites, as discrepâncias são significativas em muitos casos. Alexa corrigiu seu viés negativo com os países asiáticos, mas desta vez é Semrush quem parece ignorar os países asiáticos e árabes.
- 2) Comparando os resultados do Alexa de 2017 a 2021, esperaríamos, portanto, mudanças evolutivas relativamente menores. Não é o caso das seguintes línguas: francês, alemão, italiano e polonês, cujos números caíram de forma suspeita, uma confirmação do sentimento experimentado durante as medições de que os países europeus estavam subestimados nos números Alexa2021.

Finalmente, essas comparações tendem a confirmar situações que serão levadas em consideração ao corrigir os vieses:

- Inglês, espanhol, hindi parecem superestimados
- Francês, alemão, italiano e polonês parecem muito subestimados.
- Português e malaio parecem subestimados

Para a próxima edição, muita atenção deve ser dada a este indicador para tentar contornar a situação, talvez uma fusão de dados de serviços existentes possa ser uma alternativa para compensar vieses.

5.4 Correção de vieses

Neste ponto, não é uma questão de aplicar a correção de viés a todas as línguas no estudo, mas sim de focar apenas nas 15 principais línguas em termos de potência. Futuramente, seria interessante integrar a correção de viés ao modelo (um primeiro passo foi dado com a ponderação dos indicadores de acordo com a confiança).

Existe um método que foi usado em 2017 para produzir uma estimativa da porcentagem de conteúdo que é baseada na consistência do fator de produtividade (proporção de *conteúdo* para população conectada) para cada língua considerada e, muito importante, para o resto do línguas. Este método aplicado em 2021 leva à seguinte estimativa aproximada:

Tabela 18: Primeiro método de correção de vieses

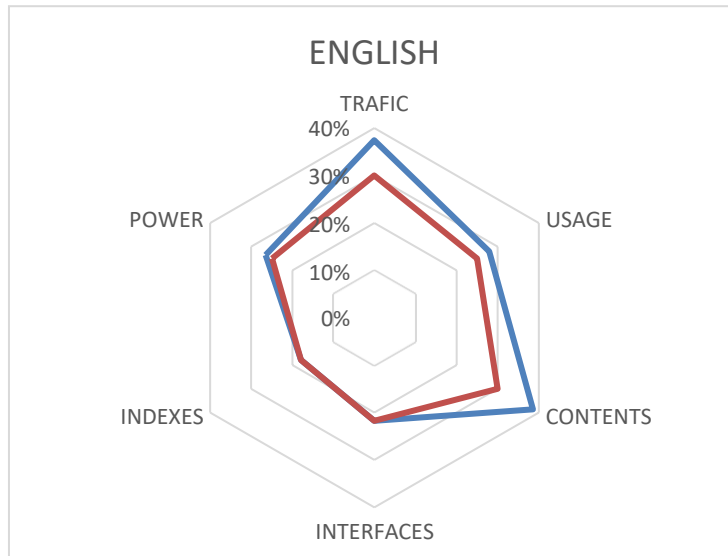
LÍNGUA	CONTEÚDO	PRODUTIVIDADE
Inglês	25,0%	1,92
Chinês	15,0%	1,02
Espanhol	7,0%	1,34
Francês	4,0%	1,55
Hindi	4,0%	0,69
Português	3,5%	1,41
Russo	3,5%	1,41
Árabe	2,5%	0,71
Alemão	2,5%	1,92
Japonês	2,5%	2,05
Malaio	1,8%	0,76
Italiano	1,4%	2,14
Turco	1,2%	1,41
Coreano	1,2%	1,53
Bengali	1,2%	0,46
Vietnamita	0,70%	0,94
RESTO	23,00%	0,58

Na ocasião, foi adicionada uma nova abordagem para correção de vieses, atuando de forma específica e direta nos respectivos vieses de cada indicador, conforme discutido nos capítulos anteriores. O padrão de pontuação da língua é revisado, indicador por indicador, à luz do que se sabe sobre viés, e um possível novo valor é registrado. A partir daí um novo valor de potência

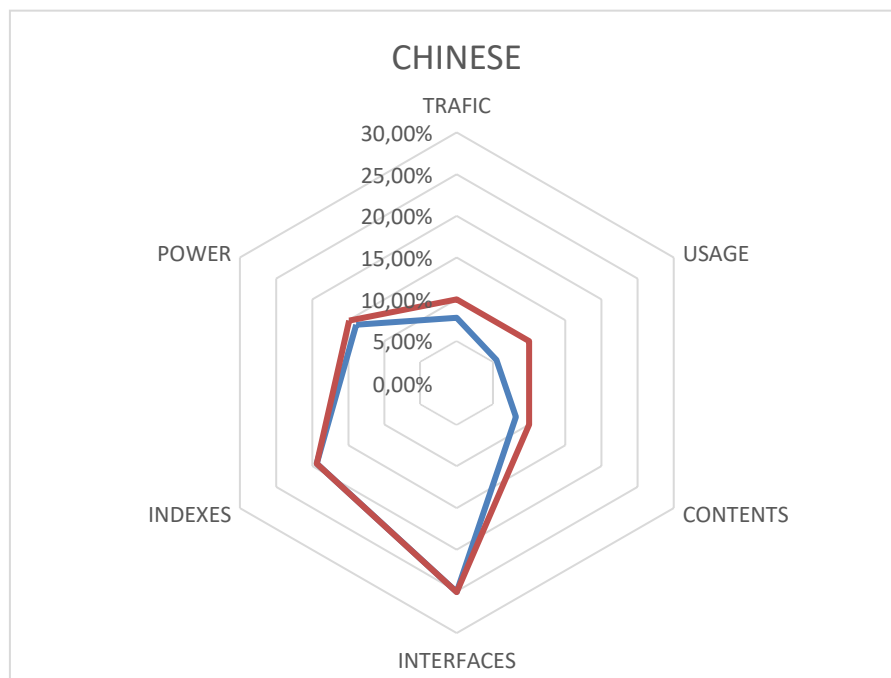
é calculado com valores arredondados. Em azul o resultado do modelo, em vermelho os valores corrigidos.

Tabela 19: Correção de vieses 2º método

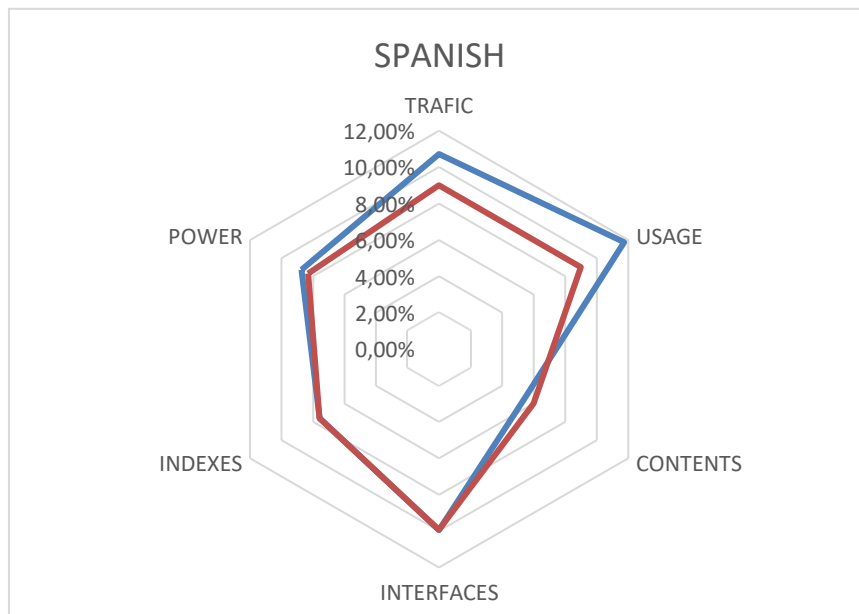
inglês	TRÁFEGO	USOS	CONTEÚDO	INTERFACES	ÍNDICE	POTÊNCIA
MODELO	37,44%	27,92%	38,61%	21,73%	17,87%	26,48%
CORREÇÃO DE VIÉS	30%	25%	30%	22%	18%	25%



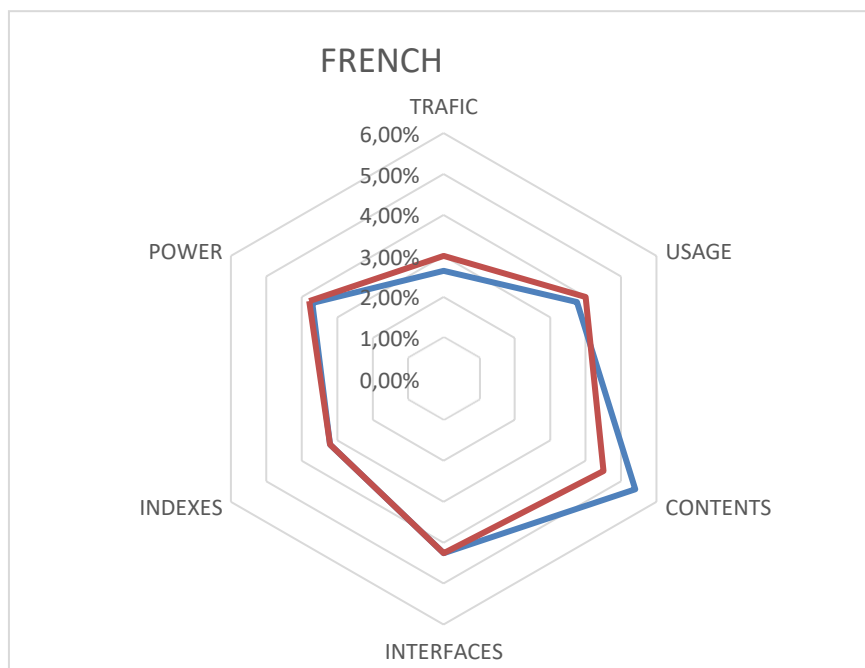
chinês	TRÁFEGO	USOS	CONTEÚDO	INTERFACES	ÍNDICE	POTÊNCIA
MODELO	7,79%	5,47%	8,18%	25,07%	19,38%	13,92%
CORREÇÃO DE VIÉS	10%	10%	10%	25%	19%	15%



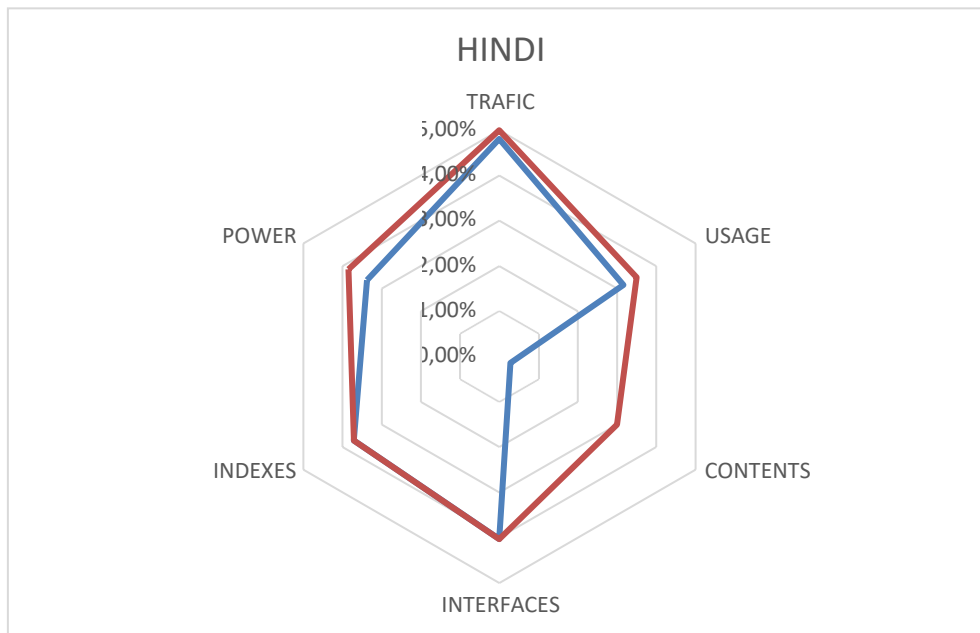
espanhol	TRÁFEGO	USOS	CONTEÚDO	INTERFACES	ÍNDICE	POTÊNCIA
MODELO	10,72%	11,74%	5,42%	9,94%	7,59%	8,73%
CORREÇÃO DE VIÉS	9%	9%	6%	dez%	8%	8%



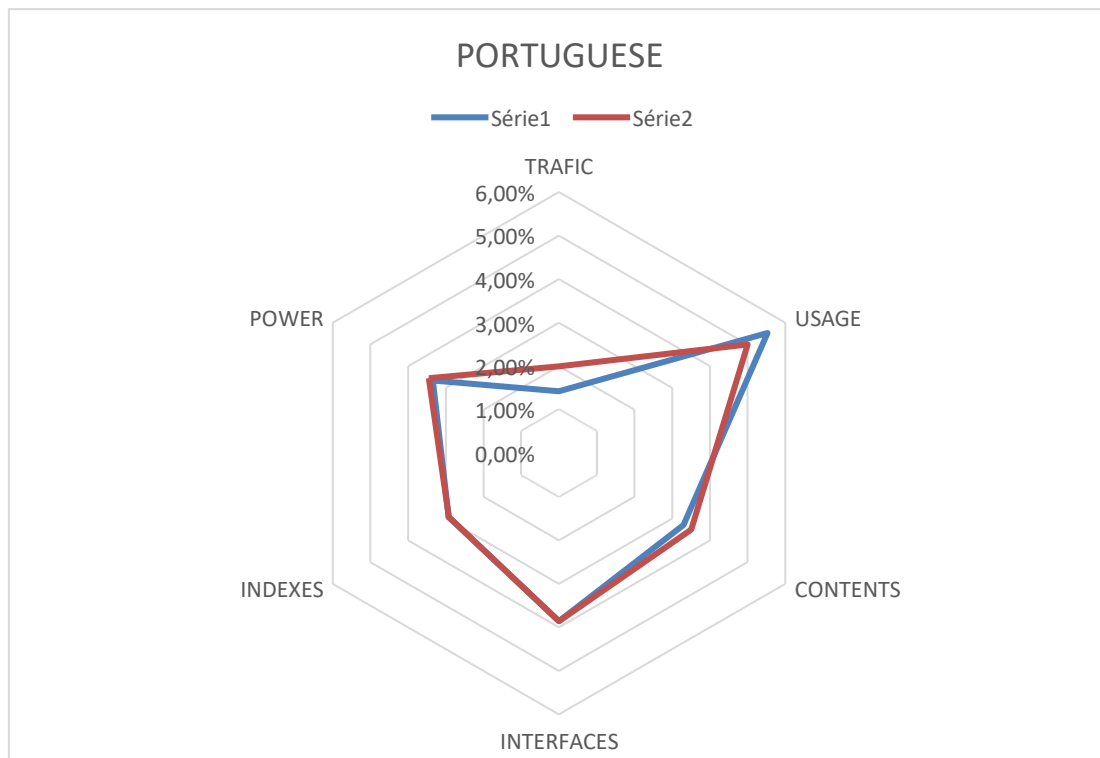
francês	TRÁFEGO	USOS	CONTEÚDO	INTERFACES	ÍNDICE	POTÊNCIA
MODELO	2,64%	3,75%	5,40%	4,26%	3,21%	3,71%
CORREÇÃO DE VIÉS	3,0%	4,0%	4,5%	4,3%	3,2%	3,8%



hindi	TRÁFEGO	USOS	CONTEÚDO	INTERFACES	ÍNDICE	POTÊNCIA
MODELO	4,81%	3,16%	0,28%	4,03%	3,71%	3,38%
CORREÇÃO DE VIÉS	5,0%	3,5%	3,0%	4,0%	3,7%	3,8%



português	TRÁFEGO	USOS	CONTEÚDO	INTERFACES	ÍNDICE	POTÊNCIA
MODELO	1,42%	5,53%	3,30%	3,85%	2,92%	3,35%
CORREÇÃO DE VIÉS	2,0%	5,5%	3%	3,9%	2,9%	3,5%



O resultado deste exercício de correção de viés é apresentado abaixo e comparado com os resultados do primeiro método de correção:

Tabela 20: Resultados da correção de vieses

	SEGUNDO MÉTODO		PRIMEIRO MÉTODO
	POTÊNCIA	CONTEÚDO	POTÊNCIA
Inglês	25%	30,0%	25%
Chinês	15%	10%	15%
Espanhol	8%	6%	7%
Francês	3,8%	4,5%	4%
Hindi	3,8%	3,0%	4%
Português	3,5%	2,8%	3,5%

É interessante notar que os resultados dos dois métodos diferentes são bastante próximos.

6. CONCLUSÕES E PERSPECTIVAS

Esta segunda versão do método de produção de indicadores de presença de línguas na Internet apresenta melhorias interessantes, nomeadamente com dados demolinguísticos mais fiáveis e na gestão de segundas línguas e multilinguismo. Está avançando, ainda, com uma abordagem consistente para estabelecer a percentagem global do número total de falantes L1+L2. Agora temos um indicador de índice mais completo. O método melhorou a análise dos vieses produzidos, aprofundando as estatísticas da Wikimédia e apresentando duas formas complementares de compensar parcialmente esses vieses.

No entanto, o método enfrenta novos desafios significativos:

- ✓ com o comportamento divergente das ferramentas de medição de *tráfego*;
- ✓ com um indicador de *conteúdo* que é muito dependente dos números da Wikimédia que não são geograficamente homogêneos, não refletem realmente a realidade do conteúdo e cuja alta sensibilidade exerce uma influência desproporcional, em particular no macroindicador gradiente;
- ✓ com um indicador de *usos* muito marcado por aplicativos de mídia social ocidentais;
- ✓ e com o fato de que a UIT não fornece mais estimativas para a percentagem *de pessoas conectadas* à Internet por país (e um problema particular sobre a percentagem exata para a Índia).

Uma nova versão está planejada antes do final de 2021, que tentará enfrentar esses desafios e tentar expandir o número de línguas processadas, empurrando o limite para línguas com mais de um milhão de falantes L1.

O objetivo da futura versão também será expandir o número de sites medidos em termos de *tráfego* para aumentar a possibilidade de resultados temáticos diferenciados mais precisos e confiáveis para determinadas línguas.

Em termos de resultados, a tendência de redução relativa da proficiência em inglês continua com uma presença estimada em cerca de 25% (ante 30% em 2017), o crescimento do chinês e o surgimento do hindi como provável quarta língua da Internet, com o francês hoje, e provavelmente acima do francês nos próximos anos.

REFERÊNCIAS

- [1] D. Pimienta, "Uma abordagem alternativa para a produção de indicadores da presença de línguas na Internet", 2020 - <http://funredes.org/lc2017/Alternativa%20Lingua%20Internet.docx>
- [2] - MAAYA, "NET.LANG: Towards a multilingual cyberspace", C&F Editions, 2012 - http://net-lang.net/lang_en
- [3] - D. Pimienta, D. Prado, Á. Blanco, "Twelve years of measuring linguistic diversity in the Internet: balance and perspectives", UNESCO, 2009 - <http://unesdoc.unesco.org/images/0018/001870/187016e.pdf>
- [4] - J. Paolillo, D. Pimienta, D. Prado, et al., "Measuring linguistic diversity on the Internet", UNESCO, 2005 - <http://www.unesco.org/new/en/communication-and-information/resources/publications-and-communication-materials/publications/full-list/measuring-linguistic-diversity-on-the-internet/>
- [5] – D. Pimienta "Indicators of Languages in the Internet", International Conference Language Technologies for All (LT4All), 4 - 6 December 2019, UNESCO, Paris <http://funredes.org/lc2019/Indicators%20Language%20Internet.pdf>

ANEXO 1. LISTA DE MICROINDICADORES E FONTES

MICRO-INDICADOR	MODELO	TEMA	FONTE DE URL
Amazon US - Valor em libra 2017	CONTEÚDOS	Book	Retomada de 2017
Valor de profundidade da Wikipédia	CONTEÚDOS	Ency	https://meta.wikimedia.org/wiki/List_of_Wikipedias
Número de usuários ativos da Wikipédia	CONTEÚDOS	Ency	https://meta.wikimedia.org/wiki/List_of_Wikipedias
Número de edições da Wikimédia	CONTEÚDOS	Ency	https://meta.wikimedia.org/wiki/List_of_Wikipedias
Número de livros wiki por língua	CONTEÚDOS	Book	https://meta.wikimedia.org/wiki/Wikibooks/Table
Número de artigos da Wikipedia por língua	CONTEÚDOS	Ency	https://meta.wikimedia.org/wiki/List_of_Wikipedias
Número de artigos WikiQuote por língua	CONTEÚDOS	Book	https://stats.wikimedia.org/wikiquote/FR/Sitemap.htm
Número de artigos WikiSource por língua	CONTEÚDOS	Book	Número de artigos WikiSource por lingua
Número de artigos da Wikiversidade por língua	CONTEÚDOS	S/T	https://stats.wikimedia.org/wikiiversity/EN/Sitemap.htm
Número de artigos do Wikcionário por língua	CONTEÚDOS	Dict	https://stats.wikimedia.org/wiktionary/EN/Sitemap.htm
Número de artigos WikiNews por língua	CONTEÚDOS	News	https://stats.wikimedia.org/wikinews/EN/Sitemap.htm
Número de artigos do WikiVoyages por língua	CONTEÚDOS	Tur	https://stats.wikimedia.org/wikivoyage/EN/Sitemap.htm
Projeção do índice T para comércio eletrônico em 2021	CONTEÚDOS	e.com	https://translated.com/les-langues-qui-comptent
Índice de administração online	ÍNDICE	S/T	https://publicadministration.un.org/egovkb/Data-Center*
Índice de participação eletrônica	ÍNDICE	S/T	https://publicadministration.un.org/egovkb/Data-Center
Índice de serviços online	ÍNDICE	Infra	https://publicadministration.un.org/egovkb/Data-Center
Índice de capital humano	ÍNDICE	ICT	https://publicadministration.un.org/egovkb/Data-Center
Índice de infraestrutura de telecomunicações	ÍNDICE	Gov	https://publicadministration.un.org/egovkb/Data-Center
Índice de preparação digital global da Cisco de 2019	ÍNDICE	S/T	https://www.cisco.com/c/dam/en_us/about/csr/reports/global-digital-readiness-index.pdf
Índice de prontidão de IA do governo 2020	ÍNDICE	ICT	https://static1.squarespace.com/static/58b2e92c1e5b6c828058484e/t/5f7747f29ca3c20ecb598f7c/1601653137399/AI+Readiness+Report.pdf
Pontuações de liberdade na Internet	ÍNDICE	Book	https://freedomhouse.org/countries/freedom-net/scores
Índice de conectividade global	ÍNDICE	Gov	https://www.huawei.com/minisite/gci/en/country-rankings.html
Índice Global de Cibersegurança 2018	ÍNDICE	Gov	https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-GCI.01-2018-PDF-E.pdf
Índice de comércio eletrônico UNCTAD B2C, 2020	ÍNDICE	Gov	https://unctad.org/system/files/official-document/tn_unctad_ict4d17_en.pdf
O índice global de dados abertos	ÍNDICE	Infra	https://index.okfn.org/place/
Ranking mundial de competitividade digital 2020	ÍNDICE	Secu	https://www.imd.org/globalassets/wcc/docs/release-2020/digital/digital_2020.pdf
Índice de preparação da Frontier Technologies	ÍNDICE	Econ	https://unctad.org/system/files/official-document/tir2020_en.pdf
Índice de inovação global	ÍNDICE	AI	https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2020.pdf
Acesso ao conhecimento básico	ÍNDICE	Econ	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Acesso a informações e comunicações	ÍNDICE	Gov	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Acesso ao ensino superior	ÍNDICE	Gov	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Acesso à eletricidade (% hab.)	ÍNDICE	Infra	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Acesso à educação de qualidade (0 = desigual; 4 = igual)	ÍNDICE	S/T	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Acesso à governança online (0 = baixo; 1 = alto)	ÍNDICE	Econ	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Censura da mídia (0 = frequente; 4 = raro)	ÍNDICE	Infra	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Liberdade de expressão (0 = sem liberdade; 1 = liberdade total)	ÍNDICE	Gov	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx

Universidades ponderadas pela qualidade (pontos)	ÍNDICE	e.com	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Documentos citáveis	ÍNDICE	Gov	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Mulheres com ensino superior	ÍNDICE	Econ	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Anos de ensino superior	ÍNDICE	S/T	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Línguas de tradução do Bing Translator	INTERFACE	Tra	https://www.bing.com/translator/
Línguas suportadas pela Amazon Kindle Direct Publishing	INTERFACE	Inter	https://kdp.amazon.com/en_US/help/topic/G200673300
Línguas suportadas pela Cortana	INTERFACE	Tra	https://en.wikipedia.org/wiki/Cortana
Línguas de palavras de referência suportados	INTERFACE	Inter	https://www.wordreference.com
Línguas de tradução do WordLingo	INTERFACE	Inter	http://www.worldlingo.com/en/languages/
Línguas suportadas pelo Facebook	INTERFACE	Tra	https://www.facebook.com/langue.php
Línguas de anúncio in-stream do Facebook com suporte	INTERFACE	Tra	https://www.facebook.com/business/help/267128784014981
Línguas de tradutores gratuitos	INTERFACE	Tra	http://www.free-translator.com
Línguas suportadas pelo console do Google Play	INTERFACE	Tra	https://support.google.com/googleplay/android-developer/table/4419860?hl=fr
Línguas suportadas pelo Google Cloud	INTERFACE	Inter	https://cloud.google.com/translate/docs/languages?hl=fr
Línguas suportadas pelo Google Translate	INTERFACE	Inter	https://en.wikipedia.org/wiki/Google_Translate
Línguas suportadas pelo Google Scholar para pesquisa	INTERFACE	Inter	https://scholar.google.com/scholar_settings?scifh=1&hl=fr&as_sdt=0,5#1
Línguas suportadas pelo Paralink Translator	INTERFACE	Inter	http://paralink.com
Línguas de tradução online	INTERFACE	Tra	https://www.online-translator.com/traduction
Línguas suportadas pelo tradutor Reverso	INTERFACE	Tra	https://www.reverso.net/text_translation.aspx?lang=FR
Línguas suportadas por tradução gratuita	INTERFACE	Tra	https://www.freetranslations.org
Línguas suportadas pelo Skype	INTERFACE	Inter	https://support.skype.com/en/faq/FA34781/what-languages-are-supported-in-skype
Línguas suportadas por Systran	INTERFACE	Tra	https://support.systran.net/systranlinks/faq/
163.com	TRÁFEGO	GAM	https://www.alexa.com/siteinfo
17ok.com	TRÁFEGO	?	https://www.alexa.com/siteinfo
1and1.com	TRÁFEGO	Tool	https://www.alexa.com/siteinfo
360.cn	TRÁFEGO	Secu	https://www.alexa.com/siteinfo
4shared.com	TRÁFEGO	FiSh	https://www.alexa.com/siteinfo
500px.com	TRÁFEGO	SN-Im	https://www.alexa.com/siteinfo
6.cn	TRÁFEGO	SN-Da	https://www.alexa.com/siteinfo
A2hosting.com	TRÁFEGO	Tool	https://www.alexa.com/siteinfo
Abilogic.com	TRÁFEGO	DIR	https://www.alexa.com/siteinfo
Sobre mim	TRÁFEGO	Mktg	https://www.alexa.com/siteinfo
Academia.edu	TRÁFEGO	S/T	https://www.alexa.com/siteinfo
Adam4Adam.com	TRÁFEGO	SN-Da	https://www.alexa.com/siteinfo
Adictingames.com	TRÁFEGO	Gam	https://www.alexa.com/siteinfo
adobe.com	TRÁFEGO	ICT	https://www.alexa.com/siteinfo
Adultfinder.com	TRÁFEGO	SN-Da	https://www.alexa.com/siteinfo
Aim.com	TRÁFEGO	MSG	https://www.alexa.com/siteinfo
Alexa.com	TRÁFEGO	Mktg	https://www.alexa.com/siteinfo
Aliexpress.com	TRÁFEGO	e.com	https://www.alexa.com/siteinfo
Alipay.com	TRÁFEGO	Econ	https://www.alexa.com/siteinfo
Alivedirectory.com	TRÁFEGO	DIR	https://www.alexa.com/siteinfo
Amazon.com	TRÁFEGO	Book	https://www.alexa.com/siteinfo
Amazonaws.com	TRÁFEGO	Host	https://www.alexa.com/siteinfo
Anastasiadate.com	TRÁFEGO	SN-Da	https://www.alexa.com/siteinfo
Android	TRÁFEGO	ICT	https://www.alexa.com/siteinfo
Angel.co	TRÁFEGO	SN-Fr	https://www.alexa.com/siteinfo
Anobii.com	TRÁFEGO	SEng	https://www.alexa.com/siteinfo
Answers.com	TRÁFEGO	Q/A	https://www.alexa.com/siteinfo
Aparat.com	TRÁFEGO	Vid	https://www.alexa.com/siteinfo
maçã	TRÁFEGO	ICT	https://www.alexa.com/siteinfo
Apple Music	TRÁFEGO	SN-Mu	https://www.alexa.com/siteinfo
Apple.com/Safari	TRÁFEGO	ICT	https://www.alexa.com/siteinfo
Archives.org	TRÁFEGO	Book	https://www.alexa.com/siteinfo
Archives-ouvertes.fr	TRÁFEGO	S/T	https://www.alexa.com/siteinfo
Armorgames.com	TRÁFEGO	Gam	https://www.alexa.com/siteinfo
Arvixe.com	TRÁFEGO	Host	https://www.alexa.com/siteinfo

Arxiv.org	TRAFEGO	S/T	https://www.alex.com/siteinfo
Ashleymadison.com	TRAFEGO	SN-Da	https://www.alex.com/siteinfo
Ask.com	TRAFEGO	SEng	https://www.alex.com/siteinfo
Ask.fm	TRAFEGO	Q/A	https://www.alex.com/siteinfo
Atom.io	TRAFEGO	App	https://www.alex.com/siteinfo
Avvo.com	TRAFEGO	Q/A	https://www.alex.com/siteinfo
Babytree.com	TRAFEGO	SN-Fr	https://www.alex.com/siteinfo
Badoo.com	TRAFEGO	SN-Da	https://www.alex.com/siteinfo
Baidu.com	TRAFEGO	SEng	https://www.alex.com/siteinfo
Bandcamp.com	TRAFEGO	SN-Im	https://www.alex.com/siteinfo
Bartleby.com	TRAFEGO	Book	https://www.alex.com/siteinfo
Base-search.net	TRAFEGO	S/T	https://www.alex.com/siteinfo
Bet365.com	TRAFEGO	Gam	https://www.alex.com/siteinfo
Beyond.com	TRAFEGO	e.com	https://www.alex.com/siteinfo
bilibili.com	TRAFEGO	SN-Im	https://www.alex.com/siteinfo
Bing.com	TRAFEGO	SEng	https://www.alex.com/siteinfo
Bit.ly	TRAFEGO	Tool	https://www.alex.com/siteinfo
Bitbucket.org	TRAFEGO	App	https://www.alex.com/siteinfo
Bitcoin.com	TRAFEGO	Econ	https://www.alex.com/siteinfo
Bitshare.com	TRAFEGO	FiSh	https://www.alex.com/siteinfo
Bl.uk	TRAFEGO	Book	https://www.alex.com/siteinfo
Blackle.com	TRAFEGO	SEng	https://www.alex.com/siteinfo
Blog.com	TRAFEGO	Blog	https://www.alex.com/siteinfo
Blogadda.com/	TRAFEGO	Blog	https://www.alex.com/siteinfo
Blogcatalog.com/	TRAFEGO	Blog	https://www.alex.com/siteinfo
Blogger.com	TRAFEGO	Blog	https://www.alex.com/siteinfo
Blogspot.com	TRAFEGO	Blog	https://www.alex.com/siteinfo
Bluehost.com	TRAFEGO	Host	https://www.alex.com/siteinfo
Blurtit.com	TRAFEGO	Q/A	https://www.alex.com/siteinfo
Bnf.fr	TRAFEGO	Book	https://www.alex.com/siteinfo
Bongacams.com	TRAFEGO	Porn	https://www.alex.com/siteinfo
booking.com	TRAFEGO	Tur	https://www.alex.com/siteinfo
Books.google.com	TRAFEGO	Book	https://www.alex.com/siteinfo
Box.com	TRAFEGO	App	https://www.alex.com/siteinfo
Supports.io	TRAFEGO	App	https://www.alex.com/siteinfo
Company.com	TRAFEGO	DIR	https://www.alex.com/siteinfo
Busuu.com	TRAFEGO	EDU	https://www.alex.com/siteinfo
C9.io	TRAFEGO	Cloud	https://www.alex.com/siteinfo
Cafemom.com	TRAFEGO	Port	https://www.alex.com/siteinfo
Caim.info	TRAFEGO	S/T	https://www.alex.com/siteinfo
Canva.com	TRAFEGO	Mktg	https://www.alex.com/siteinfo
Care2.com	TRAFEGO	Advo	https://www.alex.com/siteinfo
Caringbridge.org	TRAFEGO	Health	https://www.alex.com/siteinfo
Chacha.com	TRAFEGO	SEng	https://www.alex.com/siteinfo
Chaturbate.com	TRAFEGO	Porn	https://www.alex.com/siteinfo
Chrome.com	TRAFEGO	ICT	https://www.alex.com/siteinfo
Classmates.com	TRAFEGO	SN-Fr	https://www.alex.com/siteinfo
Codeanywhere.com	TRAFEGO	Cloud	https://www.alex.com/siteinfo
Codepen.io	TRAFEGO	SN-Fr	https://www.alex.com/siteinfo
Commonsensemedia.org	TRAFEGO	SN-Fr	https://www.alex.com/siteinfo
Contentful.com	TRAFEGO	APP	https://www.alex.com/siteinfo
Couchsurfing.com	TRAFEGO	Tur	https://www.alex.com/siteinfo
Coursera	TRAFEGO	MOOC	https://www.alex.com/siteinfo
Creativecommons.org	TRAFEGO	SEng	https://www.alex.com/siteinfo
Crunchyroll.com	TRAFEGO	SN-Im	https://www.alex.com/siteinfo
Csdn.net	TRAFEGO	SN-Fr	https://www.alex.com/siteinfo
Cyworld.com	TRAFEGO	SN-Fr	https://www.alex.com/siteinfo
Dailymotion.com	TRAFEGO	Vid	https://www.alex.com/siteinfo
Dart-europe.eu	TRAFEGO	S/T	https://www.alex.com/siteinfo
Daum.net	TRAFEGO	SN-Im	https://www.alex.com/siteinfo
Deezer.com	TRAFEGO	SN-mu	https://www.alex.com/siteinfo
Delicioso	TRAFEGO	SN-Fr	https://www.alex.com/siteinfo
Dépôtfiles.com	TRAFEGO	FiSh	https://www.alex.com/siteinfo
Deviantart.com	TRAFEGO	SN-Im	https://www.alex.com/siteinfo
Discordapp.com	TRAFEGO	App	https://www.alex.com/siteinfo
disneyplus.com	TRAFEGO	Film	https://www.alex.com/siteinfo
Dmoz.org	TRAFEGO	DIR	https://www.alex.com/siteinfo
Doaj.org	TRAFEGO	DIR	https://www.alex.com/siteinfo
Douban.com	TRAFEGO	SN-Im	https://www.alex.com/siteinfo
doubleclick.net	TRAFEGO	Mktg	https://www.alex.com/siteinfo
Draugiem.lv	TRAFEGO	SN-Fr	https://www.alex.com/siteinfo

Dreamhost.com	TRAFEGO	Host	https://www.alexa.com/siteinfo
Dreamwidth.org	TRAFEGO	SN-Fr	https://www.alexa.com/siteinfo
Dropbox.com	TRAFEGO	App	https://www.alexa.com/siteinfo
Drupal.org	TRAFEGO	CMS	https://www.alexa.com/siteinfo
Duckduckgo.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
DXY.cn	TRAFEGO	Health	https://www.alexa.com/siteinfo
ebay.com	TRAFEGO	e.com	https://www.alexa.com/siteinfo
Eclipse.org	TRAFEGO	SN-Fr	https://www.alexa.com/siteinfo
Edx.org	TRAFEGO	MOOC	https://www.alexa.com/siteinfo
egnyte.com	TRAFEGO	Tool	https://www.alexa.com/siteinfo
Eharmony.com	TRAFEGO	SN-Da	https://www.alexa.com/siteinfo
Etoro.com	TRAFEGO	Econ	https://www.alexa.com/siteinfo
Etsy.com	TRAFEGO	Econ	https://www.alexa.com/siteinfo
Europeana.eu	TRAFEGO	Book	https://www.alexa.com/siteinfo
Exalead.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Excite.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Experienceproject.com	TRAFEGO	Dead	https://www.alexa.com/siteinfo
Fandom.com	TRAFEGO	VC	https://www.alexa.com/siteinfo
Fetlife.com	TRAFEGO	SN-Fr	https://www.alexa.com/siteinfo
Filefactory.com	TRAFEGO	FiSh	https://www.alexa.com/siteinfo
Fileserve.com	TRAFEGO	FiSh	https://www.alexa.com/siteinfo
Filmaffinity.com	TRAFEGO	Film	https://www.alexa.com/siteinfo
Filmow.com	TRAFEGO	Film	https://www.alexa.com/siteinfo
Flickr.com	TRAFEGO	Vid	https://www.alexa.com/siteinfo
Flipboard.fr	TRAFEGO	SN-Fr	https://www.alexa.com/siteinfo
Flixster.com	TRAFEGO	Film	https://www.alexa.com/siteinfo
FNAC.com	TRAFEGO	Book	https://www.alexa.com/siteinfo
Force.com	TRAFEGO	Mktg	https://www.alexa.com/siteinfo
Fotki.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Fotolog.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Foursquare.com	TRAFEGO	Mktg	https://www.alexa.com/siteinfo
Fun-mooc.fr	TRAFEGO	MOOC	https://www.alexa.com/siteinfo
Funnyordie.com	TRAFEGO	Hum	https://www.alexa.com/siteinfo
Futurelearn.com	TRAFEGO	MOOC	https://www.alexa.com/siteinfo
G2a.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Gaiaonline.com	TRAFEGO	SN-Fr	https://www.alexa.com/siteinfo
Gameblog.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Gamefaqs.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Geni.com	TRAFEGO	Gen	https://www.alexa.com/siteinfo
Gfycat.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Ghost.org	TRAFEGO	Blog	https://www.alexa.com/siteinfo
Gigablast.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Gigasize.com	TRAFEGO	FiSh	https://www.alexa.com/siteinfo
Girlsaskguys.com	TRAFEGO	SN-Da	https://www.alexa.com/siteinfo
GitHub.com	TRAFEGO	Tool	https://www.alexa.com/siteinfo
Gmx.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Gmx.net	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Godaddy.com	TRAFEGO	Host	https://www.alexa.com/siteinfo
GOG.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Goodreads.com	TRAFEGO	Book	https://www.alexa.com/siteinfo
Google FR	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Gotinder.com	TRAFEGO	SN-Da	https://www.alexa.com/siteinfo
Gravatar.com	TRAFEGO	Mktg	https://www.alexa.com/siteinfo
Grindr.com	TRAFEGO	SN-Da	https://www.alexa.com/siteinfo
Gutenberg.org	TRAFEGO	Book	https://www.alexa.com/siteinfo
Haosou.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Hathitrust.org	TRAFEGO	Book	https://www.alexa.com/siteinfo
Hi5.com	TRAFEGO	SN-Da	https://www.alexa.com/siteinfo
Hightail.com	TRAFEGO	Tool	https://www.alexa.com/siteinfo
Hostgator.com	TRAFEGO	Tool	https://www.alexa.com/siteinfo
Hotmail.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Huanqiu.com	TRAFEGO	News	https://www.alexa.com/siteinfo
Hubpages.com	TRAFEGO	Blog	https://www.alexa.com/siteinfo
Hulu.com	TRAFEGO	Film	https://www.alexa.com/siteinfo
Hushmail.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Ibiblio.org	TRAFEGO	Book	https://www.alexa.com/siteinfo
icloud.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Icq.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
imdb.com	TRAFEGO	Film	https://www.alexa.com/siteinfo
Imgur.com	TRAFEGO	SN-Fr	https://www.alexa.com/siteinfo
Indiblogger.in	TRAFEGO	Blog	https://www.alexa.com/siteinfo

Inflenster.com	TRAFEGO	SN-Fr	https://www.alexa.com/siteinfo
HostingInmotion.com	TRAFEGO	Host	https://www.alexa.com/siteinfo
Instagram.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Iqiyi.com	TRAFEGO	Vid	https://www.alexa.com/siteinfo
Isbn.org	TRAFEGO	Book	https://www.alexa.com/siteinfo
Italki.com	TRAFEGO	EDU	https://www.alexa.com/siteinfo
Itch.io	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Jasminedirectory.com	TRAFEGO	DIR	https://www.alexa.com/siteinfo
jd.com	TRAFEGO	e.com	https://www.alexa.com/siteinfo
Jekyllrb.com	TRAFEGO	Blog	https://www.alexa.com/siteinfo
Jetbrains.com	TRAFEGO	Tool	https://www.alexa.com/siteinfo
joinclubhouse.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Joomla.com	TRAFEGO	CMS	https://www.alexa.com/siteinfo
Journalseek.net	TRAFEGO	S/T	https://www.alexa.com/siteinfo
Jstor.org	TRAFEGO	S/T	https://www.alexa.com/siteinfo
Jurn.org	TRAFEGO	S/T	https://www.alexa.com/siteinfo
Justanswer.com	TRAFEGO	Q/A	https://www.alexa.com/siteinfo
Kaixin001.com	TRAFEGO	SN-Fr	https://www.alexa.com/siteinfo
Kakao.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Kompas.com	TRAFEGO	Port	https://www.alexa.com/siteinfo
Kongregate.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Last FM	TRAFEGO	SN-mu	https://www.alexa.com/siteinfo
Library.harvard.edu	TRAFEGO	Book	https://www.alexa.com/siteinfo
Librarything.com	TRAFEGO	SN-fr	https://www.alexa.com/siteinfo
Line.me	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Linkedin.com	TRAFEGO	SN-pr	https://www.alexa.com/siteinfo
Linux.org	TRAFEGO	ICT	https://www.alexa.com/siteinfo
Liquidweb.com	TRAFEGO	Host	https://www.alexa.com/siteinfo
Live.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Livejasmin.com	TRAFEGO	Porn	https://www.alexa.com/siteinfo
Livejournal.com	TRAFEGO	Blog	https://www.alexa.com/siteinfo
Liveleak.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Logoslibrary.eu	TRAFEGO	Book	https://www.alexa.com/siteinfo
Lycos.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Mail.aol.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Mail.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Mail.google.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Mail.ru	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Mail.yandex.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Mamba.ru	TRAFEGO	SN-Da	https://www.alexa.com/siteinfo
Match.com	TRAFEGO	SN-Da	https://www.alexa.com/siteinfo
Mediafire.com	TRAFEGO	FiSh	https://www.alexa.com/siteinfo
Medium.com	TRAFEGO	Blog	https://www.alexa.com/siteinfo
Meetic.fr	TRAFEGO	SN-Da	https://www.alexa.com/siteinfo
Meetup.com	TRAFEGO	SN-pr	https://www.alexa.com/siteinfo
Mega.io	TRAFEGO	Cloud	https://www.alexa.com/siteinfo
Mendeley.com	TRAFEGO	S/T	https://www.alexa.com/siteinfo
Messenger.yahoo.com/	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Metacafe.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Metafilter.com	TRAFEGO	Q/A	https://www.alexa.com/siteinfo
Microsoft.com	TRAFEGO	ICT	https://www.alexa.com/siteinfo
Metropoles.com	TRAFEGO	News	https://www.alexa.com/siteinfo
Microsoftonline.com	TRAFEGO	ICT	https://www.alexa.com/siteinfo
Miniclip.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Mixi.jp	TRAFEGO	SN-Da	https://www.alexa.com/siteinfo
Mocospace.com	TRAFEGO	SN-Da	https://www.alexa.com/siteinfo
Moodle.org	TRAFEGO	CMS	https://www.alexa.com/siteinfo
Mouthshut.com	TRAFEGO	Mktg	https://www.alexa.com/siteinfo
Mozilla.org	TRAFEGO	ICT	https://www.alexa.com/siteinfo
Msn.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Mubi.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
monheritage.com	TRAFEGO	Gen	https://www.alexa.com/siteinfo
Mavie.com	TRAFEGO	Dead	https://www.alexa.com/siteinfo
Myshopify.com	TRAFEGO	e.com	https://www.alexa.com/siteinfo
Myspace.com	TRAFEGO	SN-mu	https://www.alexa.com/siteinfo
Napster.com	TRAFEGO	SN-mu	https://www.alexa.com/siteinfo
Naver.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Netcraft.com	TRAFEGO	Secu	https://www.alexa.com/siteinfo
Netflix.com	TRAFEGO	Film	https://www.alexa.com/siteinfo
Newgrounds.com	TRAFEGO	Port	https://www.alexa.com/siteinfo
Nicovideo.jp	TRAFEGO	Vid	https://www.alexa.com/siteinfo

Ning.com	TRAFEGO	SN-pr	https://www.alexa.com/siteinfo
Bloc-notes-plus-plus.org	TRAFEGO	Tool	https://www.alexa.com/siteinfo
Novoed.com	TRAFEGO	MOOC	https://www.alexa.com/siteinfo
Oatd.org	TRAFEGO	S/T	https://www.alexa.com/siteinfo
Odnoklassniki.ru	TRAFEGO	SN-fr	https://www.alexa.com/siteinfo
Office.com	TRAFEGO	ICT	https://www.alexa.com/siteinfo
Ok.ru	TRAFEGO	SN-fr	https://www.alexa.com/siteinfo
Okcupid.com	TRAFEGO	SN-Da	https://www.alexa.com/siteinfo
Okezone.com	TRAFEGO	Port	https://www.alexa.com/siteinfo
Oovoo.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Openclassrooms.com	TRAFEGO	MOOC	https://www.alexa.com/siteinfo
Opengrey.eu	TRAFEGO	S/T	https://www.alexa.com/siteinfo
Openlibrary.com	TRAFEGO	Book	https://www.alexa.com/siteinfo
Openoffice.org	TRAFEGO	ICT	https://www.alexa.com/siteinfo
Openthesis.org	TRAFEGO	S/T	https://www.alexa.com/siteinfo
Opera.com	TRAFEGO	ICT	https://www.alexa.com/siteinfo
Origin.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Outlook.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Panda.tv	TRAFEGO	Vid	https://www.alexa.com/siteinfo
Paypal.com	TRAFEGO	Econ	https://www.alexa.com/siteinfo
Pen.io	TRAFEGO	Blog	https://www.alexa.com/siteinfo
Periscope.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Periscope.tv	TRAFEGO	Vid	https://www.alexa.com/siteinfo
Photobucket.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Pikiran-rakyat.com	TRAFEGO	News	https://www.alexa.com/siteinfo
Pinterest.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Playstation.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Playstore.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Plurk.com	TRAFEGO	Port	https://www.alexa.com/siteinfo
Pornhub.com	TRAFEGO	Porn	https://www.alexa.com/siteinfo
Primevideo.com	TRAFEGO	Film	https://www.alexa.com/siteinfo
Protonmail.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Qq.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Question.com	TRAFEGO	Q/A	https://www.alexa.com/siteinfo
Quora.com	TRAFEGO	Q/A	https://www.alexa.com/siteinfo
Qwant.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Rapidshare.com	TRAFEGO	FiSh	https://www.alexa.com/siteinfo
Ravelry.com	TRAFEGO	SN-Fr	https://www.alexa.com/siteinfo
Reddit.com	TRAFEGO	SN-Fr	https://www.alexa.com/siteinfo
Rediff.com	TRAFEGO	Port	https://www.alexa.com/siteinfo
Rediffmail.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Renren.com	TRAFEGO	SN-Fr	https://www.alexa.com/siteinfo
Researchgate.net	TRAFEGO	S/T	https://www.alexa.com/siteinfo
RéverbNation.com	TRAFEGO	SN-mu	https://www.alexa.com/siteinfo
Roblox.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Rumble.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Rutube.ru	TRAFEGO	Vid	https://www.alexa.com/siteinfo
Salesforce.com	TRAFEGO	App	https://www.alexa.com/siteinfo
Toad.pt	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Savefrom.net	TRAFEGO	Tool	https://www.alexa.com/siteinfo
Scielo.org	TRAFEGO	S/T	https://www.alexa.com/siteinfo
Scienceopen.com	TRAFEGO	S/T	https://www.alexa.com/siteinfo
Search.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Secondlife.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Semanticscholar.org	TRAFEGO	S/T	https://www.alexa.com/siteinfo
Sharecare.com	TRAFEGO	Q/A	https://www.alexa.com/siteinfo
Similarweb.com	TRAFEGO	Mktg	https://www.alexa.com/siteinfo
Sina.com.cn	TRAFEGO	Port	https://www.alexa.com/siteinfo
Sitebuilder.com	TRAFEGO	Tool	https://www.alexa.com/siteinfo
Skype.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Skyrock.com	TRAFEGO	Blog	https://www.alexa.com/siteinfo
Slack.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Slideshare.net	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Smugmug.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Snapchat.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
so.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Socolar.com	TRAFEGO	S/T	https://www.alexa.com/siteinfo
Sogou.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
sohu.com	TRAFEGO	Port	https://www.alexa.com/siteinfo
Somech.com	TRAFEGO	DIR	https://www.alexa.com/siteinfo
Sony.com	TRAFEGO	SN-mu	https://www.alexa.com/siteinfo

Soso.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Soundcloud.com	TRAFEGO	SN-mu	https://www.alexa.com/siteinfo
Spaces.ru	TRAFEGO	SN-fr	https://www.alexa.com/siteinfo
Spip.net	TRAFEGO	CMS	https://www.alexa.com/siteinfo
Spotify.com	TRAFEGO	SN-mu	https://www.alexa.com/siteinfo
Squarespace.com	TRAFEGO	Tool	https://www.alexa.com/siteinfo
Stackexchange.com	TRAFEGO	Q/A	https://www.alexa.com/siteinfo
Stackoverflow.com	TRAFEGO	Q/A	https://www.alexa.com/siteinfo
Startpage.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Steam.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Steampowered.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Straightdope.com	TRAFEGO	Q/A	https://www.alexa.com/siteinfo
Stumbleupon.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Sublimetext.com	TRAFEGO	App	https://www.alexa.com/siteinfo
Svbtile.com	TRAFEGO	Blog	https://www.alexa.com/siteinfo
Tagged.com	TRAFEGO	SN-Da	https://www.alexa.com/siteinfo
Taobao.com	TRAFEGO	e.com	https://www.alexa.com/siteinfo
Taringa.net	TRAFEGO	SN-fr	https://www.alexa.com/siteinfo
Teamspeak.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Teamviewer.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Technorati.com	TRAFEGO	Blog	https://www.alexa.com/siteinfo
Telegram - interface	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Telegram.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Telegram.org	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Theblogchatter.com/	TRAFEGO	Blog	https://www.alexa.com/siteinfo
Theses.fr	TRAFEGO	S/T	https://www.alexa.com/siteinfo
Tianya.cn	TRAFEGO	SN-fr	https://www.alexa.com/siteinfo
Tiktok.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Tinyurl.com	TRAFEGO	Tool	https://www.alexa.com/siteinfo
Tmall.com	TRAFEGO	e.com	https://www.alexa.com/siteinfo
Trombi.com	TRAFEGO	SN-fr	https://www.alexa.com/siteinfo
Tudou.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Tuenti.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Tumblr.com	TRAFEGO	Blog	https://www.alexa.com/siteinfo
Twitch.tv	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Twoo.com	TRAFEGO	SN-Da	https://www.alexa.com/siteinfo
Typepad.com	TRAFEGO	Blog	https://www.alexa.com/siteinfo
Udacity.com	TRAFEGO	MOOC	https://www.alexa.com/siteinfo
Udemy.com	TRAFEGO	MOOC	https://www.alexa.com/siteinfo
Uploaded.net	TRAFEGO	FiSh	https://www.alexa.com/siteinfo
Uploading.com	TRAFEGO	FiSh	https://www.alexa.com/siteinfo
Veoh.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Viadeo.com	TRAFEGO	SN-pr	https://www.alexa.com/siteinfo
Viber.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Vimeo.com	TRAFEGO	Vid	https://www.alexa.com/siteinfo
Vk.com	TRAFEGO	SN-Mu	https://www.alexa.com/siteinfo
Wattpad.com	TRAFEGO	SN-fr	https://www.alexa.com/siteinfo
Wayn.com	TRAFEGO	Tur	https://www.alexa.com/siteinfo
Wdl.org	TRAFEGO	Book	https://www.alexa.com/siteinfo
Webcrawler.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Webometrics.info	TRAFEGO	Mktg	https://www.alexa.com/siteinfo
Wechat.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Weebly.com	TRAFEGO	e.com	https://www.alexa.com/siteinfo
Weheartit.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Weibo.com	TRAFEGO	Blog	https://www.alexa.com/siteinfo
Wetransfer.com	TRAFEGO	FiSh	https://www.alexa.com/siteinfo
Whatsapp.com	TRAFEGO	MSG	https://www.alexa.com/siteinfo
Wistia.com	TRAFEGO	SN-Im	https://www.alexa.com/siteinfo
Wix.com	TRAFEGO	App	https://www.alexa.com/siteinfo
Wolframalpha.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo
Wordpress.com	TRAFEGO	CMS	https://www.alexa.com/siteinfo
Worldcat.com	TRAFEGO	Book	https://www.alexa.com/siteinfo
Worldwidescience.org	TRAFEGO	S/T	https://www.alexa.com/siteinfo
Xbox.com	TRAFEGO	Gam	https://www.alexa.com/siteinfo
Xhamster.com	TRAFEGO	Porn	https://www.alexa.com/siteinfo
Xing.com	TRAFEGO	SN-pr	https://www.alexa.com/siteinfo
Xinhuanet.com	TRAFEGO	News	https://www.alexa.com/siteinfo
Xvideos.com	TRAFEGO	Porn	https://www.alexa.com/siteinfo
yahoo.com	TRAFEGO	Mail	https://www.alexa.com/siteinfo
Yammer.com	TRAFEGO	SN-pr	https://www.alexa.com/siteinfo
Yandex.com	TRAFEGO	SEng	https://www.alexa.com/siteinfo

Yelp.com	TRÁFEGO	SEng	https://www.alexa.com/siteinfo
Youku.com	TRÁFEGO	Vid	https://www.alexa.com/siteinfo
YouTube	TRÁFEGO	Vid	https://www.alexa.com/siteinfo
Yy.com	TRÁFEGO	Vid	https://www.alexa.com/siteinfo
Zhanqi.tv	TRÁFEGO	Vid	https://www.alexa.com/siteinfo
Zhihu.com	TRÁFEGO	Q/A	https://www.alexa.com/siteinfo
Zillow.com	TRÁFEGO	e.com	https://www.alexa.com/siteinfo
Zoho.com	TRÁFEGO	Mail	https://www.alexa.com/siteinfo
Zoom.us	TRÁFEGO	MSG	https://www.alexa.com/siteinfo
Zoosk.com	TRÁFEGO	SN-Da	https://www.alexa.com/siteinfo
% De usuários do FACEBOOK por país (NapoleonCat 2021)	USOS		https://napoleoncat.com/stats/
INSTAGRAM% de usuários por país (NapoleonCat 2021)	USOS		https://napoleoncat.com/stats/
MESSENGER% de usuários por país (NapoleonCat 2021)	USOS		https://napoleoncat.com/stats/
LINKEDIN% de usuários por país (NapoleonCat 2021)	USOS		https://napoleoncat.com/stats/
LinkedIn% usuário por país (ApolloTech 2021)	USOS		https://www.apollotechnical.com/linkedin-users-by-country/
% De usuários do Twitter por país (Statista 2021)	USOS		https://www.statista.com/statistics/242606/number-of-active-users-twitter-in-selected-countries/
% De usuários do IWS 2021 FACEBOOK	USOS		https://www.internetworldstats.com/stats1.htm + stats2.htm + ... stats6.htm
% Da audiência do Facebook (Statista 2021)	USOS		https://www.statista.com/statistics/268136/top-15-countries-based-on-facebook-number-users/
% De pessoas conectadas no país no YouTube (Statista 2021)	USOS		https://www.statista.com/statistics/1219589/youtube-penetration-worldwide-by-country/
Netflix% de assinantes por país (CompariTech 2020)	USOS		https://www.comparitech.com/tv-streaming/netflix-subscribers/
% De audiência do Pinterest (Statista 2021)	USOS		https://www.statista.com/statistics/328106/mercados-penetración-de-pinteres/
REDDIT% de usuários por país (Statista 2021)	USOS		https://backlinko.com/reddit-users
2012/21% dos downloads cumulativos do OpenOffice por país	USOS		http://www.openoffice.org/stats/countries.html
# Servidores de Internet seguros	USOS		https://data.worldbank.org/indicator/IT.NET.SECR
% Assinatura de banda larga fixa no país (BM 2021)	USOS		https://data.worldbank.org/indicator/IT.NET.BBND.P2
% Assinatura de telefone fixo + celular no país (BM 2021)	USOS		https://data.worldbank.org/indicator/IT.MLT.MAIN.P2 + https://data.worldbank.org/indicator/IT.CEL.SETS.P2

TIPOLOGIA	Qtde	TEMA
?	1	
Advo	1	Pressão
App	10	Aplicativo
Blog	20	Blog
Book	18	Livros
Cloud	3	Nuvem (computação)
CMS	5	Sistema de gerenciamento de conteúdo
DIR	7	diretório
e.com	9	Comércio eletrônico
Econ	5	Economia
EDU	2	Curso
FiSh	11	Compartilhar arquivos
Film	8	Filmes
Gam	20	Jogos
Gen	2	Genealogia
Health	2	Saúde
Host	7	Hospedagem na web
Hum	1	Humor

ICT	13	TIC
Mail	17	Correio eletrônico
Mktg	10	Marketing
MOOC	8	MOOC
MSG	2,3	Serviço de mensageiro
News	4	Jornais
Porn	6	Pornografia
Port	8	Portal
Q/A	13	Pergunta Resposta
S/T	22	Ciência e tecnologia (pesquisa)
Secu	2	Segurança
SEng	26	Motor de busca
SN-Da	20	Reuniões de redes sociais
SN-Fr	28	Amizade nas redes sociais
SN-Im	24	Imagens de mídia social
SN-Mu	10	Música de mídia social
SN-pr	6	Redes sociais profissionais
Tool	14	Ferramentas informáticas
Tur	3	turismo
VC	1	Comunidades virtuais
Vid	13	Vídeo

ANEXO 2: MACROLÍNGUAS

CÓDIGO ISO	MACRO LÍNGUA	NÚMERO DE LINGUAS FUNDIDO
<i>ara</i>	<i>Arabe</i>	29
<i>aym</i>	<i>Aimará</i>	2
<i>aze</i>	<i>Azerbaijani</i>	3
<i>bal</i>	<i>Baloutchi</i>	3
<i>bik</i>	<i>Bikol</i>	8
<i>bnc</i>	<i>Bontok</i>	5
<i>bua</i>	<i>Buryat</i>	3
<i>chm</i>	<i>Mari</i>	2
<i>cre</i>	<i>Cree</i>	6
<i>del</i>	<i>Delaware</i>	2
<i>den</i>	<i>Boniche</i>	2
<i>din</i>	<i>Dinka</i>	5
<i>doi</i>	<i>Dogri</i>	2
<i>est</i>	<i>Estoniano</i>	2
<i>fas</i>	<i>Persa</i>	2
<i>ful</i>	<i>Fulfulde</i>	9
<i>gba</i>	<i>Gbaya</i>	6
<i>gon</i>	<i>Gondi</i>	3
<i>grb</i>	<i>Grebo</i>	5
<i>grn</i>	<i>Guarani</i>	5
<i>hai</i>	<i>Haida</i>	2
<i>hbs</i>	<i>Servo-croata</i>	4
<i>hmn</i>	<i>Hmong</i>	25
<i>iku</i>	<i>Inuktitut</i>	2
<i>ipk</i>	<i>Inupiatun</i>	2
<i>jrb</i>	<i>Judaico-árabe</i>	5
<i>kau</i>	<i>Kanuri</i>	3
<i>kln</i>	<i>Kalenjin</i>	9
<i>kok</i>	<i>Concani</i>	2
<i>kom</i>	<i>Komis</i>	2
<i>kon</i>	<i>Kongo</i>	3
<i>kpe</i>	<i>Kpelle</i>	2
<i>kur</i>	<i>Curdo</i>	3
<i>lah</i>	<i>Lahnda</i>	7
<i>lav</i>	<i>Letão</i>	2
<i>luy</i>	<i>Luyia</i>	14
<i>man</i>	<i>Mandingo</i>	6
<i>mlg</i>	<i>Malgaxe</i>	11
<i>mon</i>	<i>Mongol</i>	3
<i>msa</i>	<i>Malaio</i>	36
<i>mwr</i>	<i>Marwari</i>	6
<i>nep</i>	<i>Nepalês</i>	2
<i>oji</i>	<i>Ojibwe</i>	7
<i>ori</i>	<i>Oriya</i>	2
<i>orm</i>	<i>Galla</i>	4
<i>pus</i>	<i>Pasto</i>	3
<i>que</i>	<i>Quechua</i>	42
<i>raj</i>	<i>Rajasthan</i>	6
<i>rom</i>	<i>Romani</i>	6
<i>sqi</i>	<i>Albanês</i>	4
<i>srd</i>	<i>Da sardenha</i>	4
<i>swa</i>	<i>Suaíli</i>	2
<i>syr</i>	<i>Siriaco</i>	2
<i>tmh</i>	<i>Tamasheq</i>	4
<i>uzb</i>	<i>Uzbeque</i>	2
<i>yid</i>	<i>Iídiche</i>	2
<i>zap</i>	<i>Zapoteca</i>	57
<i>zha</i>	<i>Zhuang</i>	16
<i>zho</i>	<i>Chinês</i>	15
<i>zza</i>	<i>Dimli</i>	2

ANEXO 3: LISTA DE PAÍSES OU TERRITÓRIOS DE QUE A UIT NÃO FORNECE DADOS

Código ISO	NOME DO PAÍS	POPULAÇÃO
AX	Ilhas Aland	27.652
AS	Samoa Americana	55.990
IO	Território Britânico do Oceano Índico	4.000
BQ	Caraíbas holandesas	18.740
CX	Ilha do Natal	1.170
CC	Ilhas Cocos (Keeling)	630
CK	Ilhas Cook	15.000
CW	Curaçao	140.000
GF	Guiana Francesa	366.590
GP	Guadalupe	454.800
GU	Guam	139.550
IM	Ilha de Man	88.085
MQ	Martinica	377.100
NF	Ilha Norfolk	1.500
KP	<i>Coréia do Norte</i>	<i>25.579.000</i>
MP	Ilhas Marianas do Norte	53.280
PW	Palau	17.550
PN	Pitcairn	36
RE	Reunião	751.580
BL	San Bartolome	7.850
MF	São Martim	28.500
PM	São Pedro e Miquelão	6.340
SX	São Martim	33.470
TC	Ilhas Turcas e Caicos	30.170
VA	<i>Estado do Vaticano</i>	<i>330</i>
EH	<i>Saara Ocidental</i>	<i>544.150</i>
	TOTAL	28.689.463

Existem duas razões possíveis pelas quais o país ou território foi excluído dos dados da UIT:

- 1) É um território cujos dados estão incluídos em um determinado país.
- 2) Não há fonte ou estimativa do percentual de pessoas conectadas à Internet (em itálico na tabela).

ANEXO 4: RESULTADOS PARA TODAS AS LÍNGUAS

Classificação			.Conect.L.M.	Pop.M.	TRAFF.	Conect.L.	USOD	CONT.	INTERF.	ÍNDICE	PANELA.	CAPAC.	GRAD..
	ISO	Total ou média - -->	100%	100%	100%	54,7%	100%	100%	100%	100%	100%	0,75	0,74
		Resto	10,13%	12,6%	7,90%	43,76%	8,59%	2,88%	0,02%	6,91%	6,07%	0,48	0,60
54	afr	AFRIKAANS	0,19%	0,17%	0,08%	59,75%	0,11%	0,15%	0,10%	0,17%	0,13%	0,79	0,73
102	aka	AKAN	0,06%	0,09%	0,02%	38,80%	0,05%	0,00%	0,01%	0,05%	0,03%	0,35	0,49
60	amh	AMARICO	0,21%	0,55%	0,09%	20,57%	0,11%	0,01%	0,12%	0,11%	0,11%	0,19	0,51
8	ara	ÁRABE	3,89%	3,53%	2,30%	60,14%	3,02%	2,05%	4,29%	3,01%	3,09%	0,88	0,80
74	asm	ASSAMÊS	0,11%	0,15%	0,12%	40,03%	0,08%	0,00%	0,03%	0,09%	0,07%	0,49	0,66
119	awa	AWADHI	0,03%	0,04%	0,03%	39,25%	0,02%	0,00%	0,00%	0,03%	0,02%	0,43	0,60
42	aze	AZERBAIJANI	0,31%	0,23%	0,26%	74,76%	0,16%	0,11%	0,17%	0,27%	0,22%	0,94	0,69
106	bal	BALUTCHI	0,05%	0,09%	0,06%	30,72%	0,04%	0,00%	0,00%	0,03%	0,03%	0,36	0,63
127	bam	BAMANANKAN	0,03%	0,14%	0,01%	12,94%	0,02%	0,00%	0,00%	0,01%	0,01%	0,10	0,42
53	bar	BÁVARO	0,22%	0,14%	0,10%	87,68%	0,17%	0,00%	0,00%	0,33%	0,14%	0,97	0,61
94	bel	BIELO-RÚSSIA	0,06%	0,04%	0,02%	82,27%	0,03%	0,03%	0,03%	0,06%	0,04%	1,00	0,66
15	ben	BENGALI	1,14%	2,58%	1,22%	24,15%	1,13%	0,26%	0,72%	0,84%	0,88%	0,34	0,78
112	bew	BETAWI	0,04%	0,05%	0,01%	47,69%	0,05%	0,00%	0,00%	0,04%	0,02%	0,50	0,57
34	bho	BHOJPURI	0,37%	0,51%	0,40%	39,85%	0,27%	0,00%	0,03%	0,32%	0,23%	0,46	0,63
118	bik	BIKOL	0,03%	0,04%	0,01%	43,03%	0,04%	0,00%	0,00%	0,03%	0,02%	0,51	0,65
109	bjj	KANAUJI	0,04%	0,06%	0,05%	40,00%	0,03%	0,00%	0,00%	0,04%	0,03%	0,45	0,62
116	bug	BUGIS	0,04%	0,04%	0,01%	47,94%	0,04%	0,00%	0,00%	0,03%	0,02%	0,50	0,57
63	bul	BÚLGARO	0,10%	0,08%	0,05%	70,34%	0,08%	0,13%	0,08%	0,12%	0,09%	1,18	0,92
69	ceb	CEBUANO	0,12%	0,15%	0,06%	43,15%	0,19%	0,00%	0,02%	0,11%	0,08%	0,54	0,69
38	ces	TCHECO	0,19%	0,13%	0,07%	81,37%	0,13%	0,50%	0,18%	0,25%	0,22%	1,70	1,14
55	dan	DINAMARQUÊS	0,10%	0,05%	0,04%	97,82%	0,08%	0,26%	0,08%	0,16%	0,12%	2,19	1,22
9	deu	ALEMÃO	2,09%	1,30%	1,32%	87,65%	1,95%	5,84%	2,97%	2,98%	2,86%	2,19	1,37
123	doi	DOGRI	0,03%	0,04%	0,03%	40,00%	0,02%	0,00%	0,00%	0,02%	0,02%	0,46	0,63
107	dyu	JULA	0,07%	0,12%	0,02%	30,85%	0,04%	0,00%	0,00%	0,04%	0,03%	0,24	0,43
37	ell	GREGO	0,18%	0,13%	0,21%	77,71%	0,17%	0,37%	0,19%	0,24%	0,22%	1,75	1,23
1	eng	INGLÊS	15,30%	13,01%	37,4%	64,33%	27,9%	38,61%	21,73%	17,87%	26,48%	2,04	1,73
125	ewe	OVELHA	0,03%	0,05%	0,01%	31,78%	0,02%	0,00%	0,00%	0,02%	0,01%	0,26	0,45
19	fas	PERSA	0,95%	0,81%	0,55%	64,58%	0,39%	0,74%	0,75%	0,81%	0,70%	0,87	0,73
44	fin	FINLANDÊS	0,09%	0,06%	0,04%	89,67%	0,06%	0,74%	0,08%	0,14%	0,19%	3,42	2,09
4	fra	FRANCÊS	3,00%	2,58%	2,64%	63,67%	3,75%	5,40%	4,26%	3,21%	3,71%	1,44	1,24
70	ful	FULUILDE	0,19%	0,31%	0,07%	33,16%	0,09%	0,00%	0,00%	0,12%	0,08%	0,25	0,42
89	grn	GUARANI	0,08%	0,06%	0,03%	68,83%	0,06%	0,00%	0,01%	0,07%	0,04%	0,64	0,51
73	gsw	SUÍÇO ALEMÃO	0,10%	0,06%	0,08%	91,56%	0,09%	0,00%	0,01%	0,17%	0,08%	1,21	0,72
28	guj	GUZERATE	0,44%	0,60%	0,53%	40,49%	0,35%	0,05%	0,24%	0,39%	0,34%	0,56	0,76
91	hat	CRIOULO HAITI	0,05%	0,08%	0,06%	38,59%	0,06%	0,00%	0,03%	0,03%	0,04%	0,50	0,70
45	hau	HAOUSSA	0,43%	0,72%	0,16%	32,61%	0,16%	0,00%	0,10%	0,28%	0,19%	0,26	0,44
20	hbs	SERVO-CROATA	0,27%	0,19%	0,14%	77,78%	0,21%	2,49%	0,22%	0,31%	0,61%	3,14	2,21
26	heb	HEBRAICO	0,14%	0,09%	0,08%	85,46%	0,11%	2,20%	0,13%	0,19%	0,47%	5,24	3,35
103	hil	HILIGAYNON	0,05%	0,06%	0,02%	43,08%	0,07%	0,00%	0,00%	0,04%	0,03%	0,51	0,65
5	hin	HINDI	4,26%	5,80%	4,81%	40,18%	3,16%	0,28%	4,03%	3,71%	3,38%	0,58	0,79
82	hmn	HMONG	0,09%	0,07%	0,06%	64,80%	0,05%	0,00%	0,03%	0,09%	0,05%	0,72	0,61
75	hne	CHHATTISGARHI	0,12%	0,16%	0,13%	40,00%	0,08%	0,00%	0,00%	0,10%	0,07%	0,45	0,62
41	hun	HÚNGARO	0,18%	0,12%	0,08%	79,92%	0,15%	0,57%	0,13%	0,20%	0,22%	1,79	1,22

83	hye	ARMÊNIO	0,05%	0,04%	0,02%	69,86%	0,03%	0,14%	0,02%	0,05%	0,05%	1,41	1,11
101	ibb	IBIBIO	0,08%	0,10%	0,03%	41,98%	0,03%	0,00%	0,00%	0,06%	0,03%	0,31	0,41
62	ibo	IBO	0,22%	0,28%	0,08%	42,02%	0,08%	0,00%	0,05%	0,16%	0,10%	0,35	0,45
97	ilo	ILOCANO	0,05%	0,06%	0,03%	43,82%	0,08%	0,00%	0,00%	0,05%	0,03%	0,56	0,69
12	ita	ITALIANO	0,91%	0,66%	0,51%	75,65%	0,97%	3,39%	1,22%	1,20%	1,37%	2,09	1,51
27	jav	JAVANÊS	0,58%	0,66%	0,20%	47,74%	0,69%	0,00%	0,14%	0,51%	0,35%	0,53	0,61
10	jpn	JAPONÊS	2,07%	1,22%	1,98%	92,62%	1,76%	3,55%	2,77%	3,01%	2,52%	2,07	1,22
93	kab	AMAZIGH	0,07%	0,07%	0,04%	62,12%	0,05%	0,00%	0,00%	0,06%	0,04%	0,58	0,51
30	kan	CANARIM	0,42%	0,57%	0,47%	40,12%	0,31%	0,08%	0,23%	0,36%	0,31%	0,55	0,75
104	kas	CASHMERE	0,05%	0,07%	0,06%	38,84%	0,04%	0,00%	0,00%	0,04%	0,03%	0,45	0,63
110	kau	KANURI	0,06%	0,09%	0,02%	39,21%	0,02%	0,00%	0,00%	0,04%	0,02%	0,29	0,40
56	kaz	CAZAQUE	0,18%	0,13%	0,07%	76,98%	0,10%	0,07%	0,10%	0,17%	0,11%	0,90	0,64
64	khm	KHMER	0,14%	0,17%	0,07%	43,40%	0,16%	0,02%	0,08%	0,09%	0,09%	0,53	0,66
121	kik	GIKUYU	0,03%	0,08%	0,01%	22,57%	0,03%	0,00%	0,01%	0,03%	0,02%	0,22	0,53
111	kin	KINYARWANDA	0,06%	0,13%	0,02%	24,69%	0,02%	0,00%	0,01%	0,04%	0,02%	0,19	0,42
132	kln	KALENJIN	0,02%	0,04%	0,01%	22,62%	0,02%	0,00%	0,00%	0,01%	0,01%	0,21	0,50
137	kmb	QUIMBUNDO	0,00%	0,02%	0,00%	16,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,14	0,48
108	kok	CONCANI	0,04%	0,06%	0,05%	39,76%	0,03%	0,00%	0,00%	0,04%	0,03%	0,46	0,63
130	kon	KONGO	0,02%	0,12%	0,01%	11,62%	0,02%	0,00%	0,00%	0,01%	0,01%	0,09	0,44
14	kor	COREANO	0,93%	0,79%	0,93%	64,73%	0,99%	0,85%	1,10%	0,95%	0,96%	1,22	1,03
136	ktu	KITUBA	0,01%	0,05%	0,00%	10,00%	0,01%	0,00%	0,00%	0,00%	0,00%	0,07	0,39
40	kur	CURDO	0,32%	0,24%	0,20%	73,02%	0,28%	0,04%	0,15%	0,29%	0,22%	0,89	0,67
39	lah	LAHNDA	0,31%	0,96%	0,41%	17,43%	0,26%	0,01%	0,15%	0,18%	0,22%	0,23	0,71
134	lua	LUBA-KASAĪ	0,01%	0,07%	0,00%	10,05%	0,01%	0,00%	0,00%	0,00%	0,01%	0,07	0,40
117	lug	GANDA	0,05%	0,11%	0,01%	25,01%	0,02%	0,00%	0,00%	0,03%	0,02%	0,18	0,39
133	luy	LUYIA	0,01%	0,03%	0,00%	22,98%	0,01%	0,00%	0,00%	0,01%	0,01%	0,20	0,48
95	mad	MADURO	0,07%	0,08%	0,02%	47,70%	0,08%	0,00%	0,00%	0,06%	0,04%	0,50	0,57
75	mag	MAGAHİ	0,15%	0,20%	0,16%	39,99%	0,11%	0,00%	0,00%	0,13%	0,09%	0,45	0,62
51	mai	MAITHILI	0,24%	0,33%	0,25%	39,28%	0,18%	0,00%	0,02%	0,20%	0,15%	0,44	0,62
35	mal	MALAIALA	0,28%	0,37%	0,35%	42,54%	0,26%	0,04%	0,18%	0,25%	0,23%	0,62	0,80
120	man	MANDINGO	0,04%	0,08%	0,01%	26,96%	0,03%	0,00%	0,00%	0,02%	0,02%	0,20	0,42
2,3	mar	MARATI	0,70%	0,96%	0,79%	40,06%	0,52%	0,06%	0,44%	0,61%	0,52%	0,54	0,74
99	mey	HASSANIYYA	0,07%	0,09%	0,03%	43,68%	0,05%	0,00%	0,00%	0,05%	0,03%	0,35	0,44
77	mlg	MALGAXE	0,03%	0,18%	0,01%	9,79%	0,03%	0,32%	0,01%	0,01%	0,07%	0,40	2,21
92	mon	MONGOL	0,06%	0,06%	0,03%	58,99%	0,04%	0,01%	0,02%	0,06%	0,04%	0,65	0,61
126	mos	MÒORÉ	0,03%	0,08%	0,01%	23,19%	0,02%	0,00%	0,00%	0,02%	0,01%	0,18	0,42
11	msa	MALAIIO	2,20%	2,36%	0,89%	51,00%	2,79%	0,79%	1,91%	1,99%	1,76%	0,75	0,80
67	mwr	MARWARI	0,14%	0,20%	0,16%	39,81%	0,11%	0,00%	0,00%	0,13%	0,09%	0,45	0,62
52	mya	BIRMANÊS	0,24%	0,41%	0,08%	31,85%	0,25%	0,03%	0,11%	0,14%	0,14%	0,35	0,60
86	nap	NAPOLETANO-C	0,07%	0,06%	0,03%	74,39%	0,08%	0,00%	0,00%	0,10%	0,05%	0,84	0,62
58	nep	NEPALÊS	0,16%	0,25%	0,09%	35,70%	0,14%	0,03%	0,14%	0,11%	0,11%	0,45	0,69
22	nld	HOLANDÊS	0,40%	0,24%	0,19%	92,02%	0,42%	1,13%	0,47%	0,60%	0,53%	2,26	1,34
90	nod	THAI N.	0,07%	0,06%	0,03%	66,47%	0,08%	0,00%	0,00%	0,07%	0,04%	0,70	0,57
122	nya	CHICHEWA	0,04%	0,14%	0,01%	15,87%	0,02%	0,00%	0,01%	0,02%	0,02%	0,12	0,42
43	ori	ORIYA	0,30%	0,41%	0,33%	39,96%	0,22%	0,01%	0,14%	0,26%	0,21%	0,51	0,70
84	orm	OROMO	0,13%	0,36%	0,04%	20,07%	0,06%	0,00%	0,01%	0,07%	0,05%	0,14	0,39
36	pan	PUNJABI OR.	0,33%	0,50%	0,44%	35,80%	0,30%	0,00%	0,03%	0,27%	0,23%	0,45	0,69
17	pol	POLONÊS	0,58%	0,39%	0,31%	81,17%	0,53%	1,57%	0,69%	0,73%	0,74%	1,88	1,26

6	por	PORTUGUÊS	3,05%	2,49%	1,42%	67,16%	5,53%	3,30%	3,85%	2,92%	3,35%	1,35	1,10
57	pus	PACHTOU	0,16%	0,51%	0,20%	17,49%	0,16%	0,00%	0,06%	0,09%	0,11%	0,22	0,69
85	que	QUECHUA	0,07%	0,07%	0,04%	56,82%	0,09%	0,00%	0,01%	0,07%	0,05%	0,66	0,64
78	raj	RAJASTHAN	0,11%	0,16%	0,13%	38,99%	0,08%	0,00%	0,00%	0,10%	0,07%	0,44	0,62
32	ron	ROMENA	0,32%	0,23%	0,15%	75,66%	0,26%	0,25%	0,30%	0,35%	0,27%	1,18	0,86
135	run	RUNDI	0,01%	0,11%	0,00%	4,67%	0,01%	0,00%	0,00%	0,00%	0,00%	0,04	0,42
7	rus	RUSSO	3,51%	2,49%	1,81%	77,20%	2,28%	3,38%	3,88%	3,78%	3,11%	1,25	0,88
100	sat	SANTHALI	0,05%	0,07%	0,06%	39,17%	0,04%	0,00%	0,00%	0,05%	0,03%	0,44	0,62
68	sin	CINGALÊS	0,12%	0,17%	0,06%	39,46%	0,11%	0,09%	0,05%	0,11%	0,09%	0,53	0,73
66	sik	ESLOVACO	0,11%	0,07%	0,04%	82,47%	0,07%	0,12%	0,08%	0,13%	0,09%	1,30	0,86
114	sna	SHONA	0,05%	0,09%	0,02%	30,31%	0,03%	0,00%	0,02%	0,03%	0,02%	0,26	0,46
72	snd	SINDI	0,11%	0,32%	0,15%	18,73%	0,10%	0,01%	0,03%	0,06%	0,08%	0,24	0,70
98	som	SOMALI	0,06%	0,21%	0,04%	15,24%	0,06%	0,00%	0,02%	0,03%	0,03%	0,16	0,57
79	sot	SOTO. DO SUL	0,13%	0,13%	0,06%	56,47%	0,08%	0,00%	0,01%	0,12%	0,07%	0,51	0,49
105	sou	SOUTH THAI	0,05%	0,04%	0,02%	66,68%	0,06%	0,00%	0,00%	0,05%	0,03%	0,70	0,57
3	spa	ESPAÑHOL	7,00%	5,24%	10,7%	73,08%	11,7%	5,42%	9,94%	7,59%	8,73%	1,67	1,25
80	sqi	ALBANÊS	0,08%	0,06%	0,05%	75,48%	0,08%	0,06%	0,03%	0,08%	0,06%	1,12	0,81
124	suk	SUKUMA	0,04%	0,08%	0,01%	25,00%	0,02%	0,00%	0,00%	0,02%	0,01%	0,18	0,40
47	sun	SÃO DE	0,27%	0,31%	0,09%	47,69%	0,33%	0,01%	0,06%	0,24%	0,17%	0,54	0,62
46	swa	SUAÍLI	0,32%	0,78%	0,12%	22,84%	0,21%	0,01%	0,20%	0,20%	0,18%	0,23	0,55
29	swe	SUECO	0,22%	0,13%	0,09%	93,49%	0,23%	0,87%	0,24%	0,34%	0,33%	2,61	1,53
25	tam	TÂMIL	0,62%	0,82%	0,71%	41,35%	0,51%	0,19%	0,39%	0,55%	0,50%	0,60	0,80
87	tat	TÁRTARO	0,07%	0,05%	0,03%	78,05%	0,04%	0,01%	0,03%	0,08%	0,04%	0,87	0,61
24	tel	TELUGU	0,69%	0,92%	0,80%	40,71%	0,53%	0,07%	0,38%	0,60%	0,51%	0,55	0,74
113	tgk	TADJIQUE	0,05%	0,08%	0,02%	32,22%	0,03%	0,00%	0,01%	0,03%	0,02%	0,29	0,49
33	tgl	TAGALO	0,24%	0,25%	0,33%	53,60%	0,43%	0,06%	0,15%	0,24%	0,24%	0,98	1,00
21	tha	TAILANDÊS	0,72%	0,59%	0,29%	66,85%	0,82%	0,33%	0,62%	0,67%	0,57%	0,98	0,80
129	tir	TIGRIGNA	0,03%	0,10%	0,01%	15,68%	0,02%	0,00%	0,00%	0,01%	0,01%	0,12	0,41
76	tsn	SETSWANA	0,14%	0,13%	0,06%	58,16%	0,09%	0,00%	0,01%	0,13%	0,07%	0,53	0,50
96	tso	TSONGA	0,08%	0,10%	0,03%	43,30%	0,04%	0,00%	0,01%	0,06%	0,04%	0,38	0,48
61	tts	THAI N. ESTE	0,18%	0,14%	0,07%	66,65%	0,20%	0,00%	0,00%	0,17%	0,10%	0,70	0,57
115	tuk	TURCOMANO	0,04%	0,07%	0,02%	31,48%	0,02%	0,02%	0,01%	0,02%	0,02%	0,32	0,55
13	tur	TURCO	1,21%	0,85%	1,03%	77,98%	1,59%	0,94%	1,43%	1,22%	1,24%	1,46	1,02
81	uig	UIGUR	0,12%	0,10%	0,04%	64,75%	0,03%	0,00%	0,03%	0,13%	0,06%	0,58	0,49
31	ukr	UCRANIANO	0,37%	0,32%	0,17%	63,96%	0,25%	0,26%	0,33%	0,40%	0,30%	0,92	0,79
131	umb	UMBUNDO	0,02%	0,07%	0,01%	16,00%	0,01%	0,00%	0,00%	0,01%	0,01%	0,14	0,48
18	urd	URDU	0,98%	2,22%	1,33%	24,12%	0,82%	0,03%	0,54%	0,65%	0,72%	0,33	0,74
49	uzb	UZBEQUE	0,27%	0,32%	0,10%	45,90%	0,13%	0,06%	0,13%	0,20%	0,15%	0,46	0,54
16	vie	VIETNAMITA	0,94%	0,74%	0,58%	69,04%	1,15%	0,46%	0,81%	0,83%	0,79%	1,07	0,85
128	vls	FLAMENCO O.	0,02%	0,01%	0,01%	90,43%	0,02%	0,00%	0,00%	0,03%	0,01%	1,12	0,68
88	wol	WOLOF	0,10%	0,12%	0,03%	46,09%	0,05%	0,00%	0,00%	0,07%	0,04%	0,36	0,43
59	xho	XHOSA	0,20%	0,19%	0,09%	59,96%	0,12%	0,02%	0,05%	0,19%	0,11%	0,59	0,54
50	yor	IORUBA	0,32%	0,42%	0,11%	41,74%	0,12%	0,00%	0,10%	0,23%	0,15%	0,36	0,47
71	zha	ZHUANG	0,17%	0,14%	0,06%	64,67%	0,04%	0,01%	0,00%	0,18%	0,08%	0,54	0,45
2	zho	CHINÊS	17,65%	14,72%	7,79%	65,59%	5,47%	8,18%	25,07%	19,38%	13,92%	0,95	0,79
48	zul	ZULU	0,29%	0,27%	0,13%	59,57%	0,17%	0,03%	0,09%	0,27%	0,16%	0,60	0,55

SCREENS OF THE VARIOUS TABS OF THE EXCEL IMPLEMENTED MODEL

MODEL2021New - Excel

1 ITU source revisited version 2021 : percentage of persons connected to the Internet per country

2 Differences from ITU source World Bank 2021

3 Added ISO codes ITU 2020 or 2019 source

4 Revisited column added Last ITU source (2017 or older), same as WB figure

5 2019 column modified Extrapolated

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2021 REVISITED	Not as of 2000	Source 2000
AFG Afghanistan	0,11	1,22	2,11	1,90	1,84	3,55	4,00	5,00	5,45	5,90	7,00	8,26	10,60	11,45		14,00	14,00		
ALB Albania	2,42	6,04	9,61	15,04	23,86	41,20	45,00	49,00	54,66	57,20	60,10	63,25	66,36	71,85		69,64	69,64		Albanian
DZA Algeria	4,63	5,84	7,38	9,45	10,18	11,23	12,50	14,90	18,20	22,50	29,50	38,20	42,95	47,69	49,04	60,00	60,00		
ASM American Samoa																	NP		
AND Andorra	26,84	37,61	48,94	70,87	70,04	78,53	81,00	81,00	86,43	94,00	95,90	96,91	97,93	98,87		91,57	95,00		Service de
AGO Angola	0,46	1,14	1,50	1,70	1,90	2,30	2,80	3,10	6,50	8,90	10,20	12,40	13,00	14,34		14,34	16,00		

MODEL2021New - Excel

D10 =SOMMEPROD(I\$5:H\$5;I10:H10)/\$C\$6

Country_Code 3 char	Total L1	7 231 699 136	% L1 +L2 connected	% L1+L2 connected	(L1+L2)/L1	AAA	AFG	ALB	DZA
Country_Code 2 char	TOTAL	10 361 716 756	on	on		AA	AF	AL	DZ
Country_Name	10 361 716 756	L1+L2				Remaining Coun	Afghanistan	Albania	Algeria
Total country L1+L2	10 361 716 756					31 125 283	35 992 400	2 916 176	90 919
% persons connected (ITU)	54,70					14,00	14,00	69,64	69,64
138 # persons connected	5 667 675 648				6 384 358 612	1,4328	5 038 936	2 030 908	54 551
World % connected	1,00				Number of		0,001	0,000	0
ISO_639	TOTAL OR AVERAGE	L1+L2	100,00	143,2819	Countries	L1	Number of spoken languages >		
aa0 Remain	1 311 512 714	10,13	18,13561	211	847 340 524	1,55	2 591 543	4 743 200	61 440
afr Afrikaans	17 631 780	0,19	0,24381	13	7 331 780	2,40	0	0	31 588
aka Akan	9 327 000	0,06	0,12897	7	8 327 000	1,12	0	0	0
amh Amharic	57 445 260	0,21	0,79435	13	32 345 260	1,78	0	0	0
ara Arabic Macro	366 212 980	3,89	5,06400	61	346 922 980	1,06	9 590	16 000	31 535
asm Assamese	15 326 200	0,11	0,21193	2	15 326 200	1,00	0	0	0
awa Awadhi	4 397 400	0,03	0,06081	2	4 352 000	1,01	0	0	0
aze Azerbaijani Macro	23 860 600	0,31	0,32994	20	23 860 600	1,00	0	13 000	0
bal Balochi Macro	8 817 350	0,05	0,12193	10	8 817 350	1,00	0	415 000	0
bam Bamanankan	14 181 800	0,03	0,19611	6	4 181 800	3,39	0	0	0
bar Bavarian	14 539 000	0,22	0,20105	4	14 539 000	1,00	0	0	0
bel Belarusian	3 914 980	0,06	0,05414	15	2 634 980	1,49	0	0	0
ben Bengal	267 653 920	1,14	3,70112	24	228 651 040	1,17	0	0	0
bew Betawi	5 000 000	0,04	0,06914	1	5 000 000	1,00	0	0	0
bho Bhojpuri	52 463 000	0,37	0,72546	4	52 303 000	1,00	0	0	0

MODEL2021New - Excel

F9 =G9*\$E\$6/100

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													
3													
4													
5													
6						TOTAL L1	7 231 699 136						
7						Total Connected	5 667 675 648						
8						TOTAL	100%	10 361 716 756		5 667 675 648	100,00	143,28%	54,70%
9						% POP L1+L2	POP. L1+L2	CON. SPEAKERS	POP. CONNECTED	% CONN. L1+L2	% CONN/L1	% INTERNAUTES/L1+L2	
10	aa0	Remain	12,66%	1 311 512 714	43,76%	573 975 181	10,13	18,14%	0,055				
11	afr	Afrikaans	0,17%	17 631 780	59,75%	10 534 467	0,19	0,24%	0,001				
12	aka	Akan	0,09%	9 327 000	38,80%	3 619 007	0,06	0,13%	0,000				
13	amh	Amharic	0,55%	57 445 260	20,57%	11 818 698	0,21	0,79%	0,001				
14	ara	Arabic Macro	3,53%	366 212 980	60,14%	220 250 699	3,89	5,06%	0,021				
15	asm	Assamese	0,15%	15 326 200	40,03%	6 135 458	0,11	0,21%	0,001				
16	awa	Awadhi	0,04%	4 397 400	39,25%	1 726 116	0,03	0,06%	0,000				
17	aze	Azerbaijani Macro	0,23%	23 860 600	74,76%	17 837 360	0,31	0,33%	0,002				
18	bal	Balochi Macro	0,09%	8 817 350	30,72%	2 708 306	0,05	0,12%	0,000				
19	bam	Bamanankan	0,14%	14 181 800	12,94%	1 834 463	0,03	0,20%	0,000				
20	bar	Bavarian	0,14%	14 539 000	87,68%	12 747 200	0,22	0,20%	0,001				
21	bel	Belarusian	0,04%	3 914 980	82,27%	3 220 988	0,06	0,05%	0,000				
22	ben	Bengali	2,58%	267 653 920	24,15%	64 647 201	1,14	3,70%	0,006				
23	bew	Betawi	0,05%	5 000 000	47,69%	2 384 532	0,04	0,07%	0,000				
24	bho	Bhojpuri	0,51%	52 463 000	39,85%	20 908 040	0,37	0,73%	0,002				
25	bik	Bikol Macro	0,04%	3 799 900	43,03%	1 634 968	0,03	0,05%	0,000				
26	bjj	Kanauji	0,06%	6 000 000	40,00%	2 400 000	0,04	0,08%	0,000				
27	bug	Bugis	0,04%	4 398 800	47,94%	2 108 851	0,04	0,06%	0,000				

Prêt

Taper ici pour rechercher

12:52 24/08/2021

MODEL2021New - Excel

I9 =LOC1*I19/LOC1*I1\$4

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	215	Country_Code 3 char							AAA	AFG	ALB	DZA	AND	AGO	AIA
2		Country_Code 2 char							AA	AF	AL	DZ	AD	AO	AI
3	329	Country_Name							Remaining Coun	Afghanistan	Albania	Algeria	Andorra	Angola	Anguilla
4															
5															
6															
7								215	1,00	1,00	1,00	1,00	1,00	1,00	1,00
8															
9	aa0	Remain							0,0833	0,1318	0,0211	0,3474	0,1852	0,2352	
10	afr	Afrikaans							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	
11	aka	Akan							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	
12	amh	Amharic							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	
13	ara	Arabic Macro							0,0003	0,0004	0,0000	0,3468	0,0000	0,0000	
14	asm	Assamese							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	
15	awa	Awadhi							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	
16	aze	Azerbaijani Macro							0,0000	0,0004	0,0000	0,0000	0,0000	0,0000	
17	bal	Balochi Macro							0,0000	0,0115	0,0000	0,0000	0,0000	0,0000	
18	bam	Bamanankan							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	
19	bar	Bavarian							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	
20	bel	Belarusian							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	
21	ben	Bengali							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	
22	bew	Betawi							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	
23	bho	Bhojpuri							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	
24	bik	Bikol Macro							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	
25	bjj	Kanauji							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	
26	bug	Bugis							0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	

Prêt

Taper ici pour rechercher

12:52 24/08/2021

MODEL2021New - Excel

Fichier Accueil Insertion Mise en page Formules Données Révision Affichage Développeur Aide Power Pivot Dites-nous ce que vous voulez faire Partager

C8 =SOMME(I8:H08)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		215 Country_Code 3 char							AAA	AFG	ALB	DZA	AND	AGO
2		Country_Code 2 char							AA	AF	AL	DZ	AD	AO
3		329 Country_Name							Remaining Coun	Afghanistan	Albania	Algeria	Andorra	Angola
4														
5														
6														
7			10 361 716 756											
8	ISO_639	TOTAL	10 361 716 756						31 125 283	35 992 400	2 916 176	90 919 760	162 000	39 305 6
9	aa0	Remain	1 311 512 714						2 591 543	4 743 200	61 440	31 588 600	30 000	9 245 610
10	afr	Afrikaans	17 631 780						-	-	-	-	-	-
11	aka	Akan	9 327 000						-	-	-	-	-	-
12	amh	Amharic	57 445 260						-	-	-	-	-	-
13	ara	Arabic Macro	366 212 980						9 590	18 000	-	31 535 000	-	-
14	asm	Assamese	15 326 200						-	-	-	-	-	-
15	awa	Awadhi	4 397 400						-	-	-	-	-	-
16	aze	Azerbaijani Macro	23 860 600						-	13 000	-	-	-	-
17	bal	Balochi Macro	8 817 350						-	415 000	-	-	-	-
18	bam	Bamanankan	14 181 800						-	-	-	-	-	-
19	bar	Bavarian	14 539 000						-	-	-	-	-	-
20	bel	Belarusian	3 914 980						-	-	-	-	-	-
21	ben	Bengali	267 653 920						-	-	-	-	-	-
22	bew	Betawi	5 000 000						-	-	-	-	-	-
23	bho	Bhojpuri	52 463 000						-	-	-	-	-	-
24	bik	Bikol Macro	3 799 900						-	-	-	-	-	-
25	bjj	Kanauji	6 000 000						-	-	-	-	-	-

Prêt

Taper ici pour rechercher

12:53 24/08/2021

MODEL2021New - Excel

Fichier Accueil Insertion Mise en page Formules Données Révision Affichage Développeur Aide Power Pivot Dites-nous ce que vous voulez faire Partager

C9 =(E9*F9*G9*19)/PUISSANCE('LOC1'!C9;2)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													
3													
4													
5													
6													
7													
8													
9	aa0	Remain	79368	161 416	9047	10329	202635034	524519	7209651	22198	326004	1252	2839974
10	afr	Afrikaans	7	21425	45	199	2405743	6832	99018	193	0	0	20748
11	aka	Akan	0	0	56	42	47441	3425	2177	0	0	0	0
12	amh	Amharic	0	0	34	48	364373	0	14955	0	0	0	715
13	ara	Arabic Macro	1336	33661	227	5403	60002884	174452	2433772	1551	80708	0	67319
14	asm	Assamese	0	0	174	120	285105	3429	8822	0	292	0	0
15	awa	Awadhi	0	0	6	7	23861	0	2432	0	0	0	0
16	aze	Azerbaijani Macro	170	0	34	901	7492659	34731	420677	2117	894	0	55893
17	bel	Belarusian	0	0	0	0	0	0	0	0	0	0	0
18	ben	Bengali	0	0	0	0	0	0	0	0	0	0	0
19	bew	Betawi	0	0	0	0	0	0	0	0	0	0	0
20	bho	Bhojpuri	0	0	0	0	0	0	0	0	0	0	0
21	bik	Bikol Macro	0	0	0	0	0	0	0	0	0	0	0
22	bjj	Kanauji	0	0	0	0	0	0	0	0	0	0	0

Prêt

Taper ici pour rechercher

12:54 24/08/2021

MODEL2021New - Excel

F11 =NB(U11:HO11)

Country_Code 3 char	AAA	AFG	ALB
215			
Country_Code 2 char	AA	AF	AL
Country_Name	Remaining Co	Afghanistan	Albania
Total country inhab.	7 231 699 136	28 943 123	2 854 57
% persons connected (ITU)	0,54		0,14
140 # persons connected	3 938 527 490		4 518 136
			1 988 00
VERSION WITH ALEXA2021			
COUNTRY MICRO-INDICATORS	AVG/SUM	TYPE	WEIGHT. #Val.
			SOURCE
11 INDEX E-Government Index	0,60	PN	Q 191 Gov
12 INDEX E-Participation Index	0,57	PN	Q 191 Gov
13 INDEX Online Service Index	0,57	PN	Q 191 Infra
14 INDEX Human Capital Index	0,69	PN	Q 191 Econ
15 INDEX Telecommunication Infrastructure Index	0,55	PN	Q 191 Infra
16 INDEX Cisco Global Digital Readiness Index 2019	11,81	IN	Q 141 ICT
17 INDEX Government AI Readiness Index 2020	44,06	IN	Q 174 AI
18 INDEX Internet Freedom Scores	50,98	IN	Q 65 Gov
19 INDEX Global Connectivity Index	51,43	IN	Q 79 Infra
20 INDEX Global Cybersecurity Index 2018	0,44	IN	Q 192 Secu
21 INDEX UNCTAD B2C E-commerce index, 2020	54,94	IN	Q 152 e.com
22 INDEX The Global Open Data Index	0,36	IN	Q 90 Gov
23 INDEX World Digital Competitiveness Ranking 2020	68,61	IN	Q 62 Econ
24 INDEX READINESS FOR FRONTIER TECHNOLOGIES INDEX	0,56	IN	Q 117 Econ

Q = quartile
W/Co = Weighting per Connected %

Prêt

Taper ici pour rechercher

12:55 24/08/2021

MODEL2021New - Excel

I11 =S(ESTVIDE(MIPII11);1)

Country_Code 3 char	AAA	AFG	ALB	DZA	AND	AGO	AIA
215							
Country_Code 2 char	AA	AF	AL	DZ	AD	AO	AI
Country_Name	Remaining C	Afghanistan	Albania	Algeria	Andorra	Angola	Anguilla
Number of occurrences	357	357	357	357	357	357	357
E-Government Index	191	0	1	1	1	1	0
E-Participation Index	191	0	1	1	1	1	0
Online Service Index	191	0	1	1	1	1	0
Human Capital Index	191	0	1	1	1	1	0
Telecommunication Infrastructure Index	191	0	1	1	1	1	0
Cisco Global Digital Readiness Index 2019	141	0	1	1	1	0	0
Government AI Readiness Index 2020	174	0	1	1	0	1	0
Internet Freedom Scores	65	0	0	0	0	1	0
Global Connectivity Index	79	0	0	0	1	0	0
Global Cybersecurity Index 2018	192	0	1	1	1	1	0
UNCTAD B2C E-commerce index, 2020	152	0	1	1	1	0	0
The Global Open Data Index	90	0	1	1	0	0	0
World Digital Competitiveness Ranking 2020	62	0	0	0	0	0	0
READINESS FOR FRONTIER TECHNOLOGIES INDEX	117	0	0	1	1	0	0
Global Innovation Index	131	0	0	1	1	0	0

Prêt

Taper ici pour rechercher

12:56 24/08/2021

MODEL2021New - Excel

Fichier Accueil Insertion Mise en page Formules Données Révision Affichage Développeur Aide Power Pivot Dites-nous ce que vous voulez faire Partager

I11 =SI(ESTVIDE(MIPII11);1;)

Country_Code 3 char	AAA	AFG	ALB
Country_Code 2 char	AA	AF	AL
Country_Name	Remaining C.	Afghanistan	Albania
Number of occurrences	357	357	357
E-Government Index	1	0	0
E-Participation Index	1	0	0
Online Service Index	1	0	0
Human Capital Index	1	0	0
Telecommunication Infrastructure Index	1	0	0
Cisco Global Digital Readiness Index 2019	1	0	0
Government AI Readiness Index 2020	1	0	0
Internet Freedom Scores	1	1	1
Global Connectivity Index	1	1	1
Global Cybersecurity Index 2018	1	0	0
UNCTAD B2C E-commerce index, 2020	1	0	0
The Global Open Data Index	1	0	0
World Digital Competitiveness Ranking 2020	1	1	1
READINESS FOR FRONTIER TECHNOLOGIES INDEX	1	1	0
Global Innovation Index	1	1	0

UIT LOC1 LOC2 PL LI MIL MIP Mp Ma EX RP1 RP2 RL RES FINAL Matrix DIFF DIFF-FR

Prêt Taper ici pour rechercher 12:56 24/08/2021

MODEL2021New - Excel

Fichier Accueil Insertion Mise en page Formules Données Révision Affichage Développeur Aide Power Pivot Dites-nous ce que vous voulez faire Partager

I11 =QUARTILE(MIPI\$111:\$HO111;N("LOC1"!!\$5>15)+N("LOC1"!!\$5>35)+N("LOC1"!!\$5>65)+N("LOC1"!!\$5>85))

Country_Code 3 char	AAA	AFG	ALB	DZA
EXTRAPOLATION	AA	AF	AL	DZ
	Remaining C.	Afghanista	Albania	Algeria
SUM/AVG	Extrap. Total	% Extrap.	Weight. Factor	Weighting type #Val
Occurrences number	3rd Quartile	1st quartile	Quart.Min.	Quart.Max
3rd Quartile	0,78	0,43	0,13	0,98
1st quartile	0,80	0,36	-	1,00
Quart.Min.	0,77	0,36	0,01	1,00
Quart.Max	0,84	0,56	-	1,00
E-Government Index	0,78	0,35	-	1,00
E-Participation Index	14,44	8,58	4,32	20,26
Online Service Index	54,03	31,34	19,07	85,48
Human Capital Index	66,00	31,00	10,00	95,00
Telecommunication Infrastructure Index	64,00	38,00	23,00	87,00
Cisco Global Digital Readiness Index 2019	0,72	0,16	0,00	0,93
Government AI Readiness Index 2020	77,88	33,53	5,60	95,90
Internet Freedom Scores	0,51	0,19	0,01	0,90
Global Connectivity Index	83,75	53,25	38,00	100,00
Global Cybersecurity Index 2018	0,71	0,37	0,24	1,00
UNCTAD B2C E-commerce index, 2020	42,11	24,13	13,56	66,08
The Global Open Data Index				
World Digital Competitiveness Ranking 2020				
READINESS FOR FRONTIER TECHNOLOGIES INDEX				
Global Innovation Index				

UIT LOC1 LOC2 PL LI MIL MIP Mp Ma EX RP1 RP2 RL RES FINAL Matrix DIFF DIFF-FR

Prêt Taper ici pour rechercher 12:57 24/08/2021

MODEL2021New - Excel

Fichier Accueil Insertion Mise en page Formules Données Révision Affichage Développeur Aide Power Pivot Dites-nous ce que vous voulez faire Partager

Formule: =SOMMEPROD(MIP1\$138:SHO\$38;PLI\$19:SHO9)/MIP1\$C538

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1																			
2																			
3	77	77	EXTRAPOLATION?>>>						NoEx	NoEx	NoEx	NoEx	W/Co	W/Co	NoEx	W/Co	Q	W/Co	W/Co
4																			
5																			
6			TYPE >>>						USAGES	USAGES	USAGES	USAGES	USAGES	USAGES	USAGES	USAGES	USAGES	USAGES	USAGES
7			WEIGHTING	AVG	RedAVG	AVG/RedAVG			%L1/pay	%L1/pay	%L1/pay	%L1/pay	%L1/pay	%L1/pay	%L1/pay	%L1/pay	%L1/pay	%L1/pay	%L1/pays
8			TOTAL						1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
9	aa0	Remain		0,0653	0,0617	1,0589			0,1080	0,0852	0,1142	0,0751	0,0738	0,0753	0,1140	0,1009	0,0656	0,0376	0,0774
10	afr	Afrikaans		0,0008	0,0007	1,1854			0,0012	0,0006	0,0014	0,0017	0,0014	0,0007	0,0013	0,0016	0,0003	0,0004	0,0013
11	aka	Akan		0,0002	0,0002	1,1498			0,0007	0,0004	0,0007	0,0006	0,0004	0,0002	0,0007	0,0006	0,0000	0,0001	0,0005
12	amh	Amharic		0,0010	0,0007	1,4457			0,0012	0,0003	0,0015	0,0006	0,0015	0,0009	0,0013	0,0018	0,0002	0,0005	0,0016
13	ara	Arabic Macro		0,0230	0,0180	1,2793			0,0338	0,0300	0,0344	0,0211	0,0232	0,0325	0,0387	0,0330	0,0385	0,0071	0,0280
14	asm	Assamese		0,0012	0,0009	1,3196			0,0012	0,0011	0,0012	0,0009	0,0008	0,0004	0,0013	0,0010	0,0000	0,0001	0,0008
15	awa	Awadhi		0,0003	0,0002	1,3020			0,0004	0,0003	0,0004	0,0002	0,0002	0,0001	0,0004	0,0003	0,0000	0,0000	0,0002
16	aze	Azerbaijani Macro		0,0034	0,0020	1,6900			0,0015	0,0024	0,0009	0,0006	0,0021	0,0013	0,0033	0,0027	0,0003	0,0004	0,0022
17	bal	Balochi Macro		0,0006	0,0004	1,2598			0,0005	0,0004	0,0005	0,0004	0,0004	0,0002	0,0007	0,0005	0,0011	0,0001	0,0003

MODEL2021New - Excel

Fichier Accueil Insertion Mise en page Formules Données Révision Affichage Développeur Aide Power Pivot Dites-nous ce que vous voulez faire Partager

Formule: =SOMMEPROD(BA\$9:BA\$146;'LOC1'!\$D\$9:\$D\$146)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1				E- Government Index	E- Participation Index	Online Service Index	Human Capital Index	Telecommunication Infrastructure Index	Cisco Global Digital Readiness Index 2019	Government AI Readiness Index 2020	Internet Freedom Scores	Global Connectivity Index	Global Cybersecurity Index 2018	UNCTAD B2C E-commerce Index, 2020	The Global Open Data Index	
2				PN	PN	PN	PN	PN	IN	IN	IN	PN	PN	PN	F	
3																
4																
5																
6																
7				WEIGHTING FACTOR	68,98	77,31	75,59	71,03	60,38	1216,88	5481,94	4630,19	4901,29	68,94	6188,57	41,7
8				TOTAL	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
9	aa0	Remain		0,0777	0,0732	0,0760	0,0835	0,0730	0,0779	0,0728	0,0695	0,0518	0,0771	0,0700	0,036	
10	afr	Afrikaans		0,0019	0,0018	0,0018	0,0019	0,0018	0,0017	0,0016	0,0028	0,0016	0,0017	0,0017	0,001	
11	aka	Akan		0,0005	0,0005	0,0005	0,0005	0,0006	0,0006	0,0004	0,0007	0,0004	0,0004	0,0005	0,000	
12	amh	Amharic		0,0008	0,0009	0,0010	0,0010	0,0004	0,0011	0,0010	0,0013	0,0010	0,0009	0,0009	0,001	
13	ara	Arabic Macro		0,0299	0,0226	0,0251	0,0329	0,0325	0,0326	0,0284	0,0333	0,0338	0,0291	0,0262	0,031	
14	asm	Assamese		0,0009	0,0012	0,0012	0,0009	0,0006	0,0008	0,0011	0,0012	0,0008	0,0011	0,0010	0,001	
15	awa	Awadhi		0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0003	0,0003	0,0002	0,0003	0,0003	0,000	
16	aze	Azerbaijani Macro		0,0030	0,0023	0,0026	0,0033	0,0033	0,0030	0,0025	0,0020	0,0041	0,0028	0,0033	0,002	
17	bal	Balochi Macro		0,0003	0,0003	0,0004	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,000	
18	bam	Bamanankan		0,0001	0,0001	0,0001	0,0001	0,0002	0,0001	0,0002	0,0001	0,0002	0,0000	0,0001	0,000	
19	bar	Bavarian		0,0028	0,0026	0,0026	0,0029	0,0032	0,0032	0,0030	0,0043	0,0031	0,0027	0,0033	0,002	
20	bel	Belarusian		0,0007	0,0006	0,0005	0,0007	0,0008	0,0006	0,0005	0,0005	0,0005	0,0005	0,0007	0,000	
21	ben	Bengali		0,0091	0,0101	0,0107	0,0093	0,0069	0,0084	0,0088	0,0113	0,0077	0,0100	0,0079	0,005	

MODEL2021New - Excel

Fichier Accueil Insertion Mise en page Formules Données Révision Affichage Développeur Aide Power Pivot Dites-nous ce que vous voulez faire

C10 = (MILIC10)/MILIC57

		MILIC10	MILIC5	MILIC7	MILIC8	MILIC9	MILIC10	MILIC11	MILIC12	MILIC13	MILIC14	MILIC15	MILIC16	MILIC17
1														
2	MIL WEIGHT	5,60	1,0000	0,5000	-	-	-	0,5000	-	0,1000	0,1000	0,1000	0,1000	0,1000
3	URL Source>>>		Retaken from	https://meta.wikir	https://meta	https://meta	https://meta	https://meta	https://meta	https://stats	Number of W	https://stats	https://stats	https://s
4		CONTENTS	CONTENTS	CONTENTS	CONTENTS	CONTENTS	CONTENTS	CONTENTS	CONTENTS	CONTENTS	CONTENTS	CONTENTS	CONTENTS	CONTE
7	TOTAL		1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
9	aa0 Remain	0,0793	0,0047	0,4859	0,0341	0,0677	0,0487	0,1270	0,0935	0,0316	0,0094	0,0918	0,0000	0,0000
10	afr Afrikaans	0,0000	0,0006	0,0024	0,0007	0,0008	0,0006	0,0017	0,0008	-	-	0,0007	-	-
11	aka Akan	0,0000	-	0,0030	0,0001	0,0000	0,0003	0,0000	-	-	-	-	-	-
12	amh Amharic	0,0000	-	0,0018	0,0002	0,0001	-	0,0003	-	-	-	0,0000	-	-
13	ara Arabic Macro	0,0013	0,0010	0,0122	0,0178	0,0200	0,0162	0,0429	0,0065	0,0078	-	0,0022	0,0000	0,0000
14	asm Assamese	0,0000	-	0,0093	0,0004	0,0001	0,0003	0,0002	-	0,0000	-	-	-	-
15	awa Awadhi	0,0000	-	0,0003	0,0000	0,0000	-	0,0000	-	-	-	-	-	-
16	aze Azerbaijani Macro	0,0002	-	0,0018	0,0030	0,0025	0,0032	0,0074	0,0089	0,0001	-	0,0018	-	-
17	bal Balochi Macro	-	-	-	-	-	-	-	-	-	-	-	-	-

Prêt

Taper ici pour rechercher

12:58 24/08/2021

MODEL2021New - Excel

Fichier Accueil Insertion Mise en page Formules Données Révision Affichage Développeur Aide Power Pivot Dites-nous ce que vous voulez faire

D9 = MOYENNE.REDUITE("RP1":Y9:M29;0,2)

		INDEX	TRAFFIC	TRAFFIC Norm.	USAGES	USAGES Norm.	CONTENTS	CONTENTS Norm.	INTERFACES	INTERFACES Norm.
1										
2										
3	77	77								
4	INDICATOR									
5										
6	NB OF MICRO-INDICATORS>>	27	340	12	14	18				
7										70,40534098
8	TOTAL	1,0000	0,7800	1,0000	0,9662	1,0000	1,0000	1,0000	0,2802	1,0000
9	aa0 Remain	0,0691	0,0617	0,0790	0,0830	0,0859	0,0288	0,0288	0,0016	0,0002
10	afr Afrikaans	0,0017	0,0007	0,0008	0,0011	0,0011	0,0015	0,0015	0,3889	0,0010
11	aka Akan	0,0005	0,0002	0,0002	0,0004	0,0005	0,0000	0,0000	0,0556	0,0001
12	amh Amharic	0,0011	0,0007	0,0009	0,0010	0,0011	0,0001	0,0001	0,3889	0,0012
13	ara Arabic Macro	0,0301	0,0180	0,0230	0,0291	0,0302	0,0205	0,0205	0,7778	0,0429
14	asm Assamese	0,0009	0,0009	0,0012	0,0008	0,0008	0,0000	0,0000	0,1667	0,0003
15	awa Awadhi	0,0003	0,0002	0,0003	0,0002	0,0002	0,0000	0,0000	0,0000	0,0000
16	aze Azerbaijani Macro	0,0027	0,0020	0,0026	0,0016	0,0016	0,0011	0,0011	0,3889	0,0017
17	bal Balochi Macro	0,0003	0,0004	0,0006	0,0004	0,0004	0,0000	0,0000	0,0556	0,0000
18	bam Bamanankan	0,0001	0,0001	0,0001	0,0002	0,0002	0,0000	0,0000	0,0000	0,0000
19	bar Bavarian	0,0033	0,0008	0,0010	0,0016	0,0017	0,0000	0,0000	0,0000	0,0000

Prêt

Taper ici pour rechercher

13:00 24/08/2021

MODEL2021New - Excel

Connexion

Fichier Accueil Insertion Mise en page Formules Données Révision Affichage Développeur Aide Power Pivot Dites-nous ce que vous voulez faire Partager

O10 =M10/O10

FINAL RESULTS **138. langues**

INDICATOR'S WEIGHT	16	11	12	8	19	18	84					
RANK	%	%	%	%	%	%	%	%	%			
POWER L1+L2	INTERNAUTE L1+L2	WORLD POP L1+L2	TRAFIC L1+L2	CONN. SPEAKERS	USAGE L1+L2	CONTENTS L1+L2	INTERFACES L1+L2	INDEXES L1+L2	POWER L1+L2	CAPACITY L1+L2	GRADIENT L1+L2	
TOTAL	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,7351	0,7363	
aa0	Remain	0,1013	0,1266	0,0790	43,76%	0,0859	0,0288	0,0002	0,0691	0,0595	0,4702	0,59
54	afr Afrikaans	0,0019	0,0017	0,0008	59,75%	0,0011	0,0015	0,0010	0,0017	0,0014	0,8045	0,74
101	aka Akan	0,0006	0,0009	0,0002	38,80%	0,0005	0,0000	0,0001	0,0005	0,0003	0,3735	0,53
58	amh Amharic	0,0021	0,0055	0,0009	20,57%	0,0011	0,0001	0,0012	0,0011	0,0012	0,2098	0,56
7	ara Arabic Mac	0,0389	0,0353	0,0230	60,14%	0,0302	0,0205	0,0429	0,0301	0,0328	0,9291	0,85
75	asm Assamese	0,0011	0,0015	0,0012	40,03%	0,0008	0,0000	0,0003	0,0009	0,0007	0,5004	0,68
119	awa Awadhi	0,0003	0,0004	0,0003	39,25%	0,0002	0,0000	0,0000	0,0003	0,0002	0,4402	0,61
38	aze Azerbaijani	0,0031	0,0023	0,0026	74,76%	0,0016	0,0011	0,0017	0,0027	0,0023	0,9797	0,72
107	bal Balochi Ma	0,0005	0,0009	0,0006	30,72%	0,0004	0,0000	0,0000	0,0003	0,0003	0,3531	0,63
128	bam Bamanankan	0,0003	0,0014	0,0001	12,94%	0,0002	0,0000	0,0000	0,0001	0,0001	0,1026	0,43
51	bar Bavarian	0,0022	0,0014	0,0010	87,68%	0,0017	0,0000	0,0000	0,0033	0,0015	1,0675	0,67
92	bel Belarusian	0,0006	0,0004	0,0002	82,27%	0,0003	0,0003	0,0003	0,0006	0,0004	1,0680	0,71
15	ben Bengali	0,0114	0,0258	0,0122	24,15%	0,0113	0,0026	0,0072	0,0084	0,0091	0,3505	0,79
112	bew Betawi	0,0004	0,0005	0,0001	47,69%	0,0005	0,0000	0,0000	0,0004	0,0003	0,5194	0,60
34	bho Bhojpuri	0,0037	0,0051	0,0040	39,85%	0,0027	0,0000	0,0003	0,0032	0,0024	0,4675	0,64
118	bik Bikol Maccr	0,0003	0,0004	0,0001	43,03%	0,0004	0,0000	0,0000	0,0003	0,0002	0,5219	0,66
111	bjj Kanauji	0,0004	0,0006	0,0005	40,00%	0,0003	0,0000	0,0000	0,0004	0,0003	0,4591	0,63

UIT LOC1 LOC2 PL LI MIL MIP Mp Ma EX RP1 RP2 RL RES FINAL Matrix DIFF DIFF-FR ...

Prêt

Taper ici pour rechercher

13:02 24/08/2021

MODEL2021New - Excel

Connexion

Fichier Accueil Insertion Mise en page Formules Données Révision Affichage Développeur Aide Power Pivot Dites-nous ce que vous voulez faire Partager

G20

MICRO INDICATOR	TYPE	THEME 1	THEME 2	Alexa rank
Amazon US - number of books 2017	CONTENT	Book		
Value of Wikimedia depth	CONTENT	Ency		
Number of active Wikipedia users	CONTENT	Ency		
Number of Wikimedia edits	CONTENT	Ency		
Number of Wiki Books per language	CONTENT	Book		
Number of wikipedia article by language	CONTENT	Ency		
WikiQuote articles per language	CONTENT	Book		
Number of WikiSource articles per language	CONTENT	Book		
Number of article Wikiversity per language	CONTENT	S/T	Science & Technology	
Number of articles Wiktionary per language	CONTENT	Dict	Dictionary	
Number of articles WikiNews per language	CONTENT	News		
Number of articles WikiVoyages per language	CONTENT	Tur		
T-Index for e-commerce Projection 2021	CONTENT	e.com		
Access to Advanced Education	INDEX	S/T		
Access to Basic Knowledge	INDEX	S/T		
Access to electricity (% of pop.)	INDEX	Infra	Infrastructure	
Access to Information and Communications	INDEX	ICT		
Access to online governance (0=low; 1=high)	INDEX	Gov		
Access to quality education (0=unequal; 4=equal)	INDEX	S/T		
Cisco Global Digital Readiness Index 2019	INDEX	ICT		
Citable documents	INDEX	Book		
E-Government Index	INDEX	Gov		
E-Participation Index	INDEX	Gov		
Freedom of expression (0=no freedom; 1=full freedom)	INDEX	Gov		
Global Connectivity Index	INDEX	Infra		
Global Cybersecurity Index 2018	INDEX	Secu	Security	

UIT LOC1 LOC2 PL LI MIL MIP Mp Ma EX RP1 RP2 RL RES FINAL Matrix DIFF DIFF-FR ...

Prêt

Taper ici pour rechercher

13:02 24/08/2021