

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 639.3.

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 demolinguística.....	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 Wikipédia.....	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

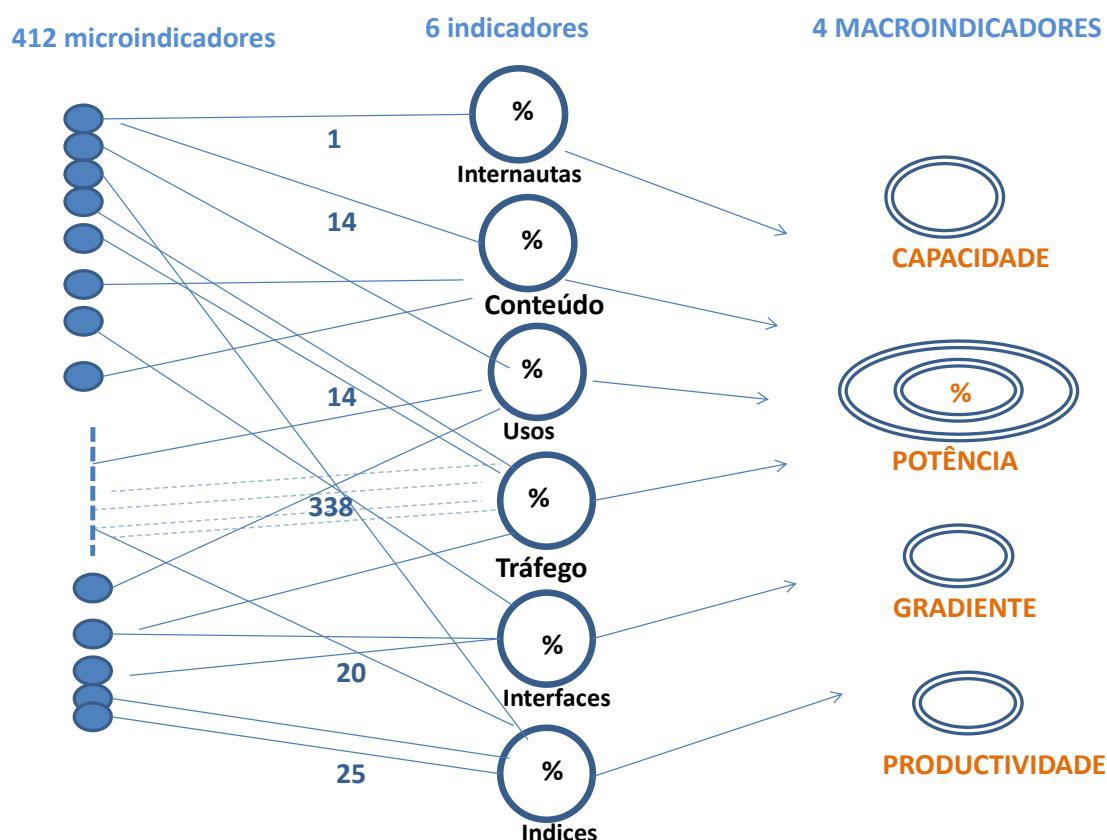
Figura 1: De microindicadores a macroindicadores.....	5
Figura 2: O processo de entrada / saída do modelo.	6

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



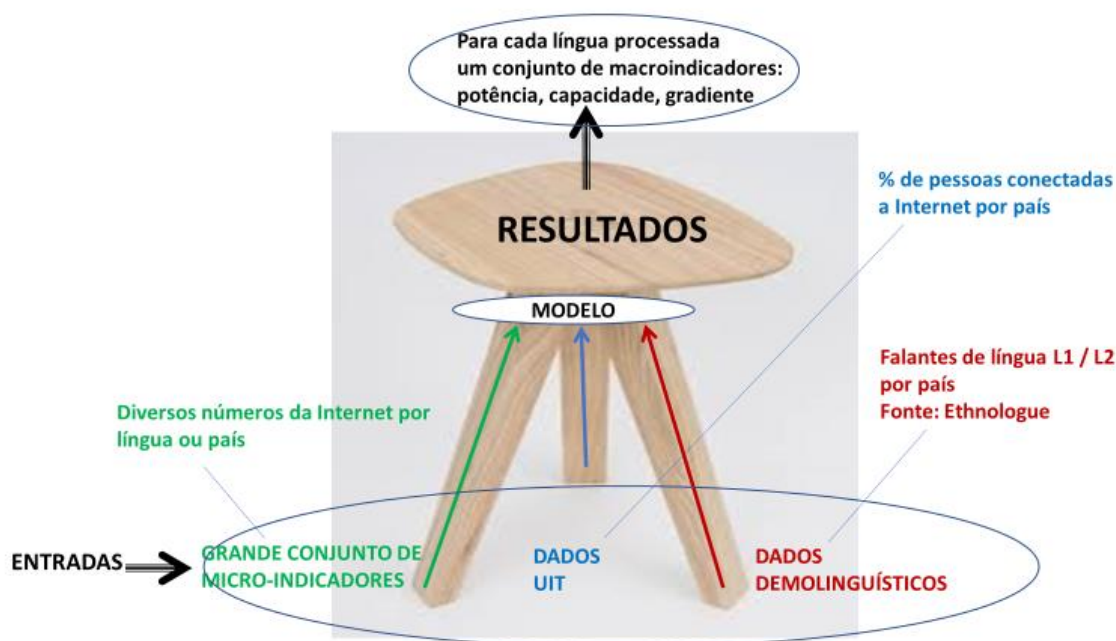
O método consiste em 3 tipos de entradas e dez sinalizadores de saída, conforme mostrado na figura a seguir.

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.

³Na 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: awhadi, 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 Wikipédia 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 profundidade¹⁹) 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 Wikipédia: *Profundidade*, definido como [Modificações / Artigos] × [Nº de artigos / Artigos] × [1 - **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 aceder 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%	19,87%	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 Wikipédia, apesar do esforço feito para combatê-lo, favorece claramente as línguas cujos falantes investiram na presença da Wikipédia. 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

		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).
TRÁFEGO	13→11	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.
INTERFACES	19	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.
USOS	12	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.

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 Wikimédia.
- 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 Wikimédia 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 Wikimédia 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%	vinde%
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

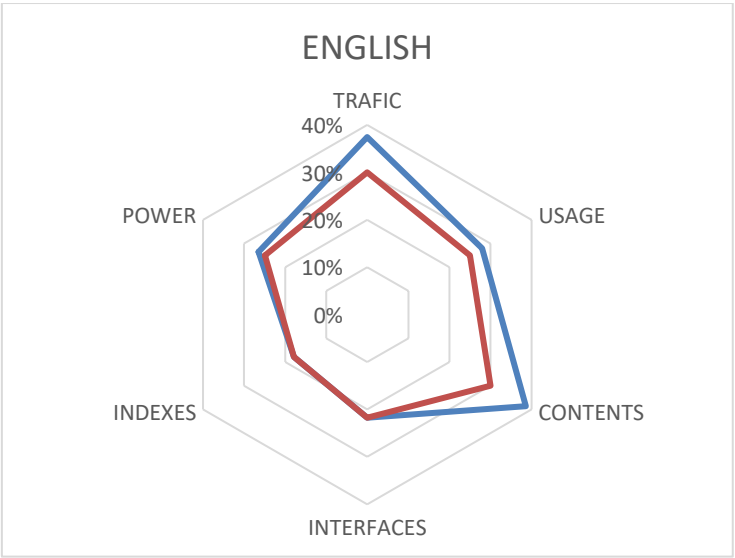
LÍNGUA	CONTEÚDO	PRODUTIVIDADE
Inglês	25,0%	1,92
Chinês	15,0%	1.02
Espanhol	7,0%	1,34
Francês	4,0%	1,55
Hindi	4,0%	0,69
Português	3,5%	1,41
Russo	3,5%	1,41
Árabe	2,5%	0,71
Alemão	2,5%	1,92
Japonês	2,5%	2.05
Malaio	1,8%	0,76
Italiano	1,4%	2,14
Turco	1,2%	1,41
Coreano	1,2%	1,53
Bengali	1,2%	0,46
Vietnamita	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

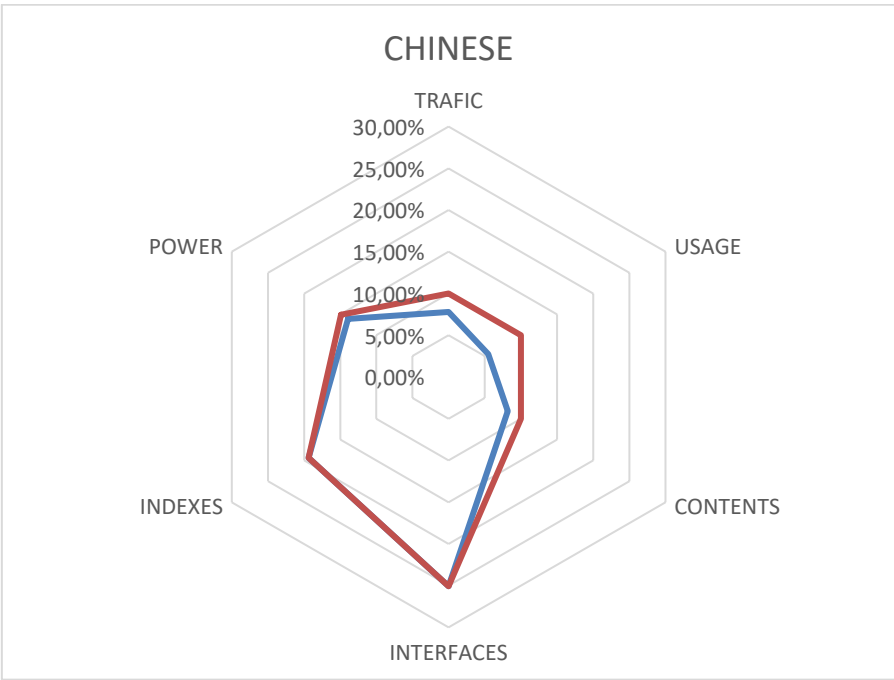
é calculado com valores arredondados. Em azul o resultado do modelo, em vermelho os valores corrigidos.

Tabela 19: Correção de vieses 2º método

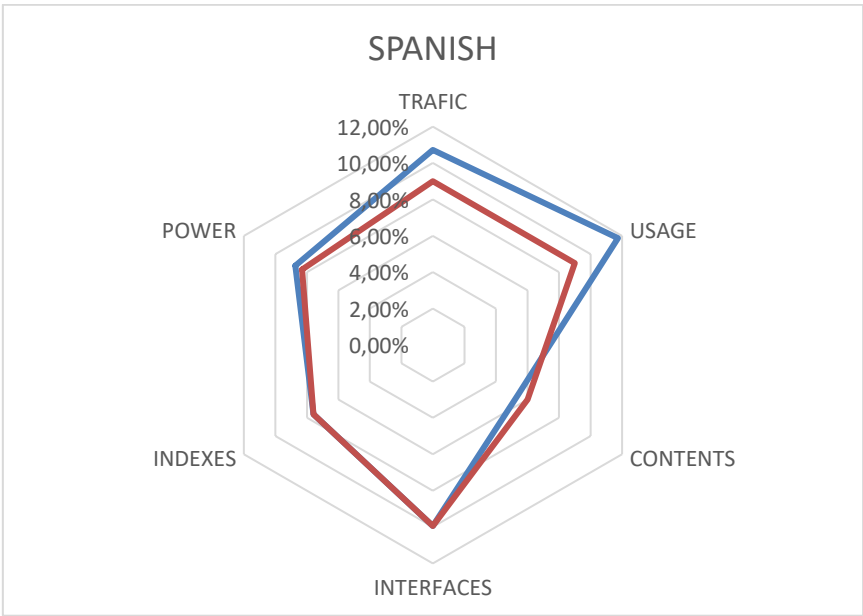
inglês	TRÁFEGO	USOS	CONTEÚDO	INTERFACES	ÍNDICE	POTÊNCIA
MODELO	37,44%	27,92%	38,61%	21,73%	17,87%	26,48%
CORREÇÃO DE VIÉS	30%	25%	30%	22%	18%	25%



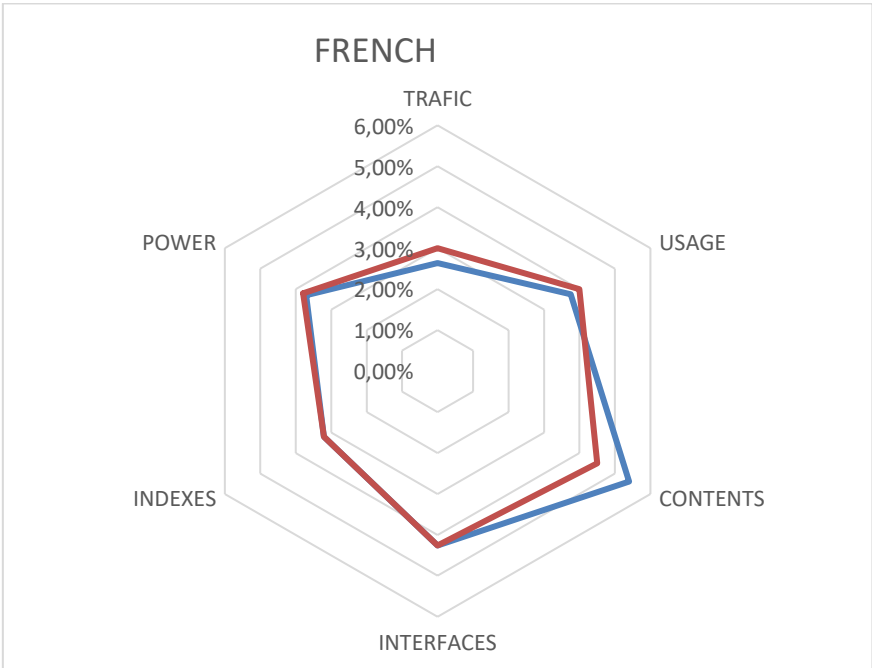
chinês	TRÁFEGO	USOS	CONTEÚDO	INTERFACES	ÍNDICE	POTÊNCIA
MODELO	7,79%	5,47%	8,18%	25,07%	19,38%	13,92%
CORREÇÃO DE VIÉS	10%	10%	10%	25%	19%	15%



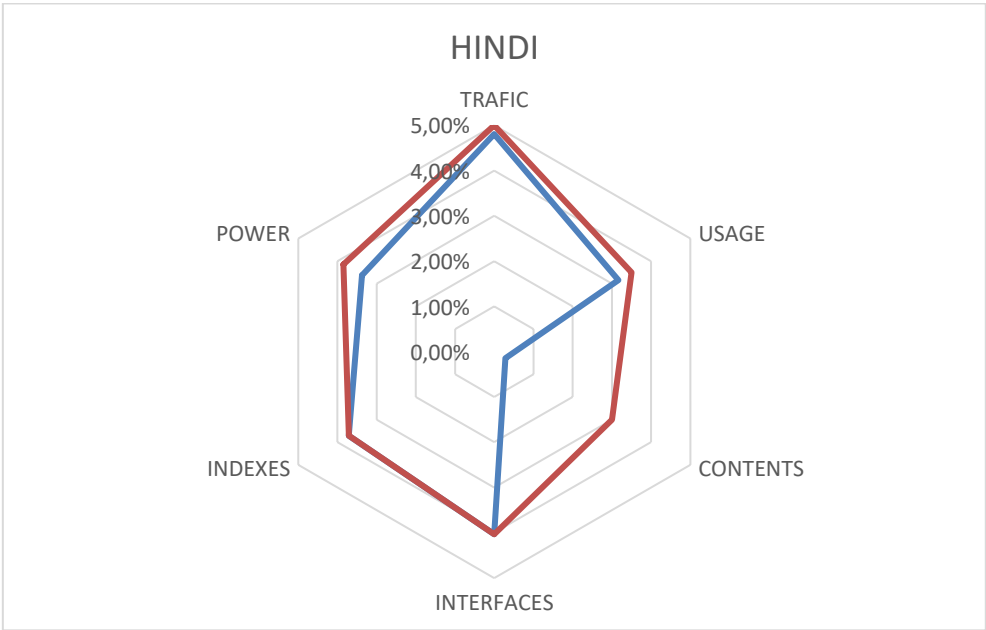
espanhol	TRÁFEGO	USOS	CONTEÚDO	INTERFACES	ÍNDICE	POTÊNCIA
MODELO	10,72%	11,74%	5,42%	9,94%	7,59%	8,73%
CORREÇÃO DE Viés	9%	9%	6%	dez%	8%	8%



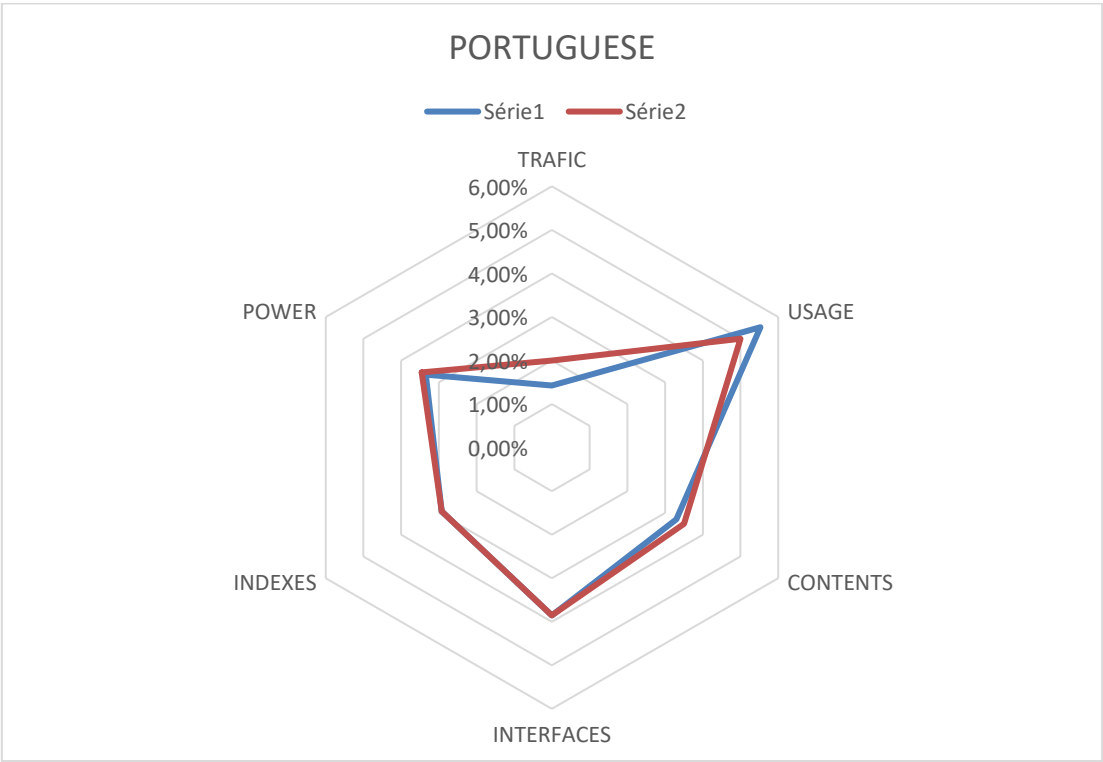
francês	TRÁFEGO	USOS	CONTEÚDO	INTERFACES	ÍNDICE	POTÊNCIA
MODELO	2,64%	3,75%	5,40%	4,26%	3,21%	3,71%
CORREÇÃO DE VIÉS	3,0%	4,0%	4,5%	4,3%	3,2%	3,8%



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/wikiversity/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
Adictinggames.com	TRÁFEGO	Gam	https://www.alexa.com/siteinfo
adobe.com	TRÁFEGO	ICT	https://www.alexa.com/siteinfo
Adultfriendfinder.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
Arvix.com	TRÁFEGO	Host	https://www.alexa.com/siteinfo

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

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

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

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

Soso.com	TRÁFEGO	SEng	https://www.alexa.com/siteinfo
Soundcloud.com	TRÁFEGO	SN-mu	https://www.alexa.com/siteinfo
Spaces.ru	TRÁFEGO	SN-fr	https://www.alexa.com/siteinfo
Spip.net	TRÁFEGO	CMS	https://www.alexa.com/siteinfo
Spotify.com	TRÁFEGO	SN-mu	https://www.alexa.com/siteinfo
Squarespace.com	TRÁFEGO	Tool	https://www.alexa.com/siteinfo
Stackexchange.com	TRÁFEGO	Q/A	https://www.alexa.com/siteinfo
Stackoverflow.com	TRÁFEGO	Q/A	https://www.alexa.com/siteinfo
Startpage.com	TRÁFEGO	SEng	https://www.alexa.com/siteinfo
Steam.com	TRÁFEGO	Gam	https://www.alexa.com/siteinfo
Steampowered.com	TRÁFEGO	Gam	https://www.alexa.com/siteinfo
Straightdope.com	TRÁFEGO	Q/A	https://www.alexa.com/siteinfo
Stumbleupon.com	TRÁFEGO	SEng	https://www.alexa.com/siteinfo
Sublimetext.com	TRÁFEGO	App	https://www.alexa.com/siteinfo
Svbtile.com	TRÁFEGO	Blog	https://www.alexa.com/siteinfo
Tagged.com	TRÁFEGO	SN-Da	https://www.alexa.com/siteinfo
Taobao.com	TRÁFEGO	e.com	https://www.alexa.com/siteinfo
Taringa.net	TRÁFEGO	SN-fr	https://www.alexa.com/siteinfo
Teamspeak.com	TRÁFEGO	MSG	https://www.alexa.com/siteinfo
Teamviewer.com	TRÁFEGO	MSG	https://www.alexa.com/siteinfo
Technorati.com	TRÁFEGO	Blog	https://www.alexa.com/siteinfo
Telegram - interface	TRÁFEGO	MSG	https://www.alexa.com/siteinfo
Telegram.com	TRÁFEGO	MSG	https://www.alexa.com/siteinfo
Telegram.org	TRÁFEGO	MSG	https://www.alexa.com/siteinfo
Theblogchatter.com/	TRÁFEGO	Blog	https://www.alexa.com/siteinfo
Theses.fr	TRÁFEGO	S/T	https://www.alexa.com/siteinfo
Tianya.cn	TRÁFEGO	SN-fr	https://www.alexa.com/siteinfo
Tiktok.com	TRÁFEGO	MSG	https://www.alexa.com/siteinfo
Tinyurl.com	TRÁFEGO	Tool	https://www.alexa.com/siteinfo
Tmall.com	TRÁFEGO	e.com	https://www.alexa.com/siteinfo
Trombi.com	TRÁFEGO	SN-fr	https://www.alexa.com/siteinfo
Tudou.com	TRÁFEGO	SN-lm	https://www.alexa.com/siteinfo
Tuenti.com	TRÁFEGO	MSG	https://www.alexa.com/siteinfo
Tumblr.com	TRÁFEGO	Blog	https://www.alexa.com/siteinfo
Twitch.tv	TRÁFEGO	Gam	https://www.alexa.com/siteinfo
Twoo.com	TRÁFEGO	SN-Da	https://www.alexa.com/siteinfo
Typepad.com	TRÁFEGO	Blog	https://www.alexa.com/siteinfo
Udacity.com	TRÁFEGO	MOOC	https://www.alexa.com/siteinfo
Udemy.com	TRÁFEGO	MOOC	https://www.alexa.com/siteinfo
Uploaded.net	TRÁFEGO	FiSh	https://www.alexa.com/siteinfo
Uploading.com	TRÁFEGO	FiSh	https://www.alexa.com/siteinfo
Veoh.com	TRÁFEGO	SN-lm	https://www.alexa.com/siteinfo
Viadeo.com	TRÁFEGO	SN-pr	https://www.alexa.com/siteinfo
Viber.com	TRÁFEGO	MSG	https://www.alexa.com/siteinfo
Vimeo.com	TRÁFEGO	Vid	https://www.alexa.com/siteinfo
Vk.com	TRÁFEGO	SN-Mu	https://www.alexa.com/siteinfo
Wattpad.com	TRÁFEGO	SN-fr	https://www.alexa.com/siteinfo
Wayn.com	TRÁFEGO	Tur	https://www.alexa.com/siteinfo
Wdl.org	TRÁFEGO	Book	https://www.alexa.com/siteinfo
Webcrawler.com	TRÁFEGO	SEng	https://www.alexa.com/siteinfo
Webometrics.info	TRÁFEGO	Mktg	https://www.alexa.com/siteinfo
Wechat.com	TRÁFEGO	MSG	https://www.alexa.com/siteinfo
Weebly.com	TRÁFEGO	e.com	https://www.alexa.com/siteinfo
Weheartit.com	TRÁFEGO	SN-lm	https://www.alexa.com/siteinfo
Weibo.com	TRÁFEGO	Blog	https://www.alexa.com/siteinfo
Wetransfer.com	TRÁFEGO	FiSh	https://www.alexa.com/siteinfo
Whatsapp.com	TRÁFEGO	MSG	https://www.alexa.com/siteinfo
Wistia.com	TRÁFEGO	SN-lm	https://www.alexa.com/siteinfo
Wix.com	TRÁFEGO	App	https://www.alexa.com/siteinfo
Wolframalpha.com	TRÁFEGO	SEng	https://www.alexa.com/siteinfo
Wordpress.com	TRÁFEGO	CMS	https://www.alexa.com/siteinfo
Worldcat.com	TRÁFEGO	Book	https://www.alexa.com/siteinfo
Worldwidescience.org	TRÁFEGO	S/T	https://www.alexa.com/siteinfo
Xbox.com	TRÁFEGO	Gam	https://www.alexa.com/siteinfo
Xhamster.com	TRÁFEGO	Porn	https://www.alexa.com/siteinfo
Xing.com	TRÁFEGO	SN-pr	https://www.alexa.com/siteinfo
Xinhuanet.com	TRÁFEGO	News	https://www.alexa.com/siteinfo
Xvideos.com	TRÁFEGO	Porn	https://www.alexa.com/siteinfo
yahoo.com	TRÁFEGO	Mail	https://www.alexa.com/siteinfo
Yammer.com	TRÁFEGO	SN-pr	https://www.alexa.com/siteinfo
Yandex.com	TRÁFEGO	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>Árabe</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>Siríaco</i>	2
<i>tmh</i>	<i>Tamasheq</i>	4
<i>uzb</i>	<i>Uzbeque</i>	2
<i>vid</i>	<i>Idiche</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>Coreia 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 Martin	28.500
PM	São Pedro e Miquelão	6.340
SX	São Martin	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			.Connect..M.	Pop.M.	TRAFF.	Conneç.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	klñ	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-KASAÏ	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	MAITHİLİ	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	MALAIO	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	slk	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