

Versión nueva y mejorada de un enfoque alternativo para la producción de indicadores lingüísticos en la Internet

Daniel Pimienta

Observatorio de la diversidad lingüística y cultural en la Internet

<http://funredes.org/lc>

Agosto 2021

Créditos: Este proyecto fue posible gracias al apoyo del [Departamento de Cultura y Educación del Ministerio de Relaciones Exteriores de Brasil](#) en el marco del [Instituto Internacional de Lengua Portuguesa](#) y bajo la coordinación de la [Cátedra UNESCO de Políticas Lingüísticas para el Multilingüismo](#). También deben otorgarse créditos a Daniel Prado, quien fue el primero en tener la idea de recolectar múltiples fuentes para medir la presencia de las lenguas en la Internet, así como para transformar datos por país en datos por lengua.

Gracias: Al profesor Gilvan Müller de Oliveira por su apoyo en cuestiones lingüísticas y la coordinación con los donantes; a Álvaro Blanco por escribir programas para Excel que cambiaron radicalmente el manejo de tantas fuentes y ortografías de lenguas y países y a David Pimienta quien escribió la programación necesaria para transformar el formato Ethnologue al formato requerido para este estudio, así como para el tratamiento de las macro-lenguas.

Advertencia: El siguiente estudio es esencialmente un trabajo estadístico basado en una amplia variedad de fuentes. La adopción de una fuente importante en este tipo de trabajo implica también y lógicamente la adopción de las reglas que sustentan los datos de esa fuente. El autor no es responsable de la lista de países y territorios considerados, establecida por la UIT, una agencia de las Naciones Unidas, ni de la lista de lenguas con más de cinco millones de hablantes de L1, según Ethnologue, así como de la agrupación en macro-lenguas, adoptada por Ethnologue, de acuerdo con la norma ISO 693.3.

RESUMEN

En un contexto de escasez de datos fiables sobre el espacio lingüístico en la Internet, se enriqueció y actualizó el enfoque alternativo de 2017 para la producción de indicadores de comportamiento en la Internet de 140 lenguas con más de 5 millones de hablantes. Se presentan las mejoras de este enfoque, que se basa en la recopilación de una gran serie de micro-indicadores sobre lenguas o países en diversos espacios o aplicaciones de la Internet (o en relación a la Internet). La utilización de los últimos datos elaborados por Ethnologue permitió disponer de las cifras demo-lingüísticas más fiables y actualizadas y, además, aportar los elementos para superar uno de los principales sesgos del método, ligado al tratamiento de los hablantes de L2. Los seis indicadores de presencia de lenguas en la Internet definidos y estudiados en 2017 (*internautas, tráfico, usos, contenidos, indexes e interfaces*), y los cuatro macro-indicadores que se deducen de ellos (*potencia, contenido capacidad, gradiente y productividad*) se reproducen después de las actualizaciones para 2021 de todas las fuentes. Los resultados muestran una disminución relativa en la presencia del inglés, en torno al 25% (frente al 30% en 2017) y un claro aumento de la presencia del chino mientras que el español se consolida en la tercera posición. El francés comparte ahora el tercer lugar con el hindi con una ventaja reducida, en comparación con 2017, con un grupo de lenguas con pesos muy similares: portugués, ruso, árabe y alemán. Al igual que en la edición de 2017, todos los posibles sesgos derivados del método, hipótesis o fuentes se examinan y se propone una estimación que tiene en cuenta estos sesgos, para las principales lenguas. Está previsto un nuevo conjunto de mejoras para finales de 2021 con la posibilidad de ampliar los resultados para las 332 lenguas con más de 1 millón de hablantes L1.

Palabras clave: Lenguas, Internet, diversidad lingüística, indicadores, sesgo

Contenido

RESUMEN	2
CONTEXTO.....	5
1. INTRODUCCIÓN	6
2. DIFERENCIAS CON LA PRIMERA VERSIÓN	7
2.1 Elección de los datos de Ethnologue como fuente demo-lingüística	7
2.2 Gestión de L2 y multilingüismo.....	8
2.3 Fuente para personas conectadas a la Internet	9
2.4 Gestión de fuentes para micro-indicadores.....	10
2.4.1 INDEXES	11
2.4.2 CONTENIDOS	11
2.4.3 TRÁFICO.....	14
2.4.4 INTERFACES	15
2.4.5 USOS	15
2.5 Resumen de indicadores.....	15
3. RESULTADOS	16
4. ANÁLISIS DE LOS RESULTADOS	19
5. ANÁLISIS DE SESGOS	21
5.1 Sesgos específicos del método	21
5.2 Sesgo de selección de fuente.....	22
5.3 Sesgo de fuente.....	22
5.3.1 Sesgos de Wikimedia	24
5.3.2 Sesgos de Alexa	31
5.4 Corrección de sesgos	32
6. CONCLUSIONES Y PERSPECTIVAS.....	37
REFERENCIAS	39
ANEXO 1. LISTA DE MICRO INDICADORES Y FUENTES	40
ANEXO 2: MACRO-LENGUAS.....	49
ANEXO 3: LISTA DE PAÍSES O TERRITORIOS DONDE LA UIT NO OFRECE DATOS	50
ANEXO 4: RESULTADOS PARA TODAS LAS LENGUAS	51

LISTA DE TABLAS Y FIGURAS

TABLAS

Tabla 1 : Los 2 tipos de ponderaciones utilizados.	6
Tabla 2: Sensibilidad de las cifras de la India sobre el porcentaje de personas conectadas a la Internet	9
Tabla 3: Factores de Wikipedia y la fórmula	13
Tabla 4: Ponderación de los indicadores de contenido	13
Tabla 5: Descripción de indicadores	15
Tabla 6 : Indicadores para los 15 lenguas principales en términos de potencia.....	16
Tabla 7 : Lenguas ordenados por porcentaje de personas conectadas	17
Tabla 8 : Lenguas ordenados por capacidad	18
Tabla 9 : Lenguas ordenados por gradiente	19
Tabla 10: Presencia en Wikipedia ordenada por valores de fórmula.....	20
Tabla 11: Evaluación del sesgo por indicador	22
Tabla 12 : Indicadores macro para los 15 lenguas principales después de ponderar los indicadores ...	24
Tabla 13: Ordenado por número de artículos de Wikipedia	25
Tabla 14: Artículos de Wikipedia ordenados por fórmula.....	26
Tabla 15: Número de wikilibros	27
Tabla 16: Número de citas (WikiQuote).....	28
Tabla 17: Número de Wikisources.....	28
Tabla 18: Número de Wikiversidad.....	29
Tabla 19: Número de entradas de Wikcionario	29
Tabla 20: Número de Wikinoticias	30
Tabla 21: Número de artículos en Wikivoyages	30
Tabla 22: Comparaciones de diferentes medidas de tráfico.....	31
Tabla 23: Primer método de corrección de sesgos.....	33
Tabla 24: Corrección de sesgo 2do método	33
Tabla 25: Resultados de la corrección de sesgo.....	37

FIGURAS

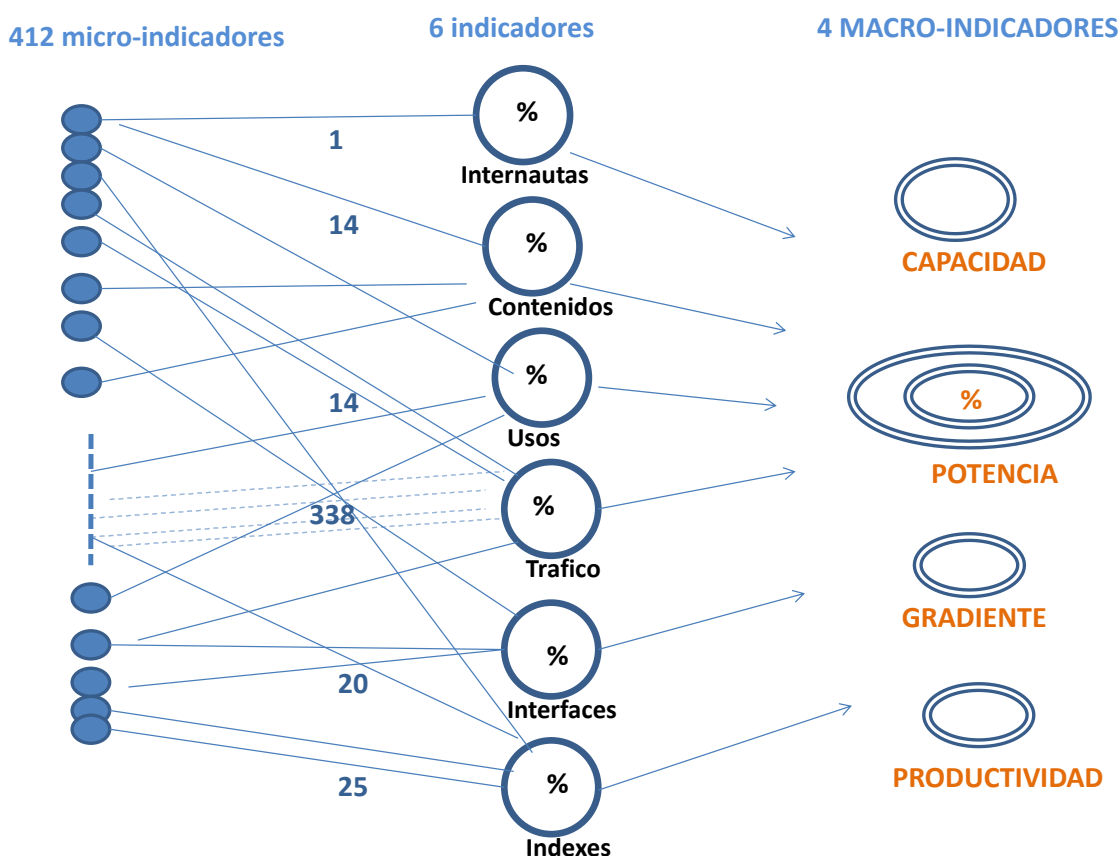
Figura 1: De micro-indicadores a macro-indicadores	5
Figura 2: El proceso de entrada / salida del modelo.	6

CONTEXTO

La primera edición de este método para producir indicadores de presencia de lenguas en la Internet se realizó en 2017 y se documentó bajo el título “Un enfoque alternativo para producir indicadores de lenguas en la Internet” ([1]) accesible en el sitio web del Observatorio en 4 versiones lingüísticas (inglés, francés, portugués y español)¹. Se invita al lector a consultar este documento antes de leer este artículo que se escribe de forma complementaria a esta primera versión, la cual presenta tanto el método como los resultados; este artículo presenta las diferencias de método y los nuevos resultados.

Como recordatorio, el método enfoca las 138 lenguas cuyo número de hablantes L1² es superior a 5 millones³ y produce indicadores para cada uno de ellas, según el siguiente diagrama (cuyas cifras se actualizan para la segunda versión).

Figura 1: De micro-indicadores a macro-indicadores



El método se compone de 3 tipos de entradas y diez indicadores de salida, como se muestra en la siguiente figura.

¹ <http://funredes.org/lc2017>

² La convención es llamar a L1 la lengua materna (o primera lengua) y L2 las segundas lenguas, entendiéndose que es necesario un dominio suficiente de una lengua para estar calificada de esta manera.

³ De hecho, el total es 128: para poder hacer comparaciones con el estudio de 2017 se han dejado 10 idiomas con menos de 5 millones de hablantes porque aparecieron en el estudio de 2017. Esos idiomas son: awjadi, bielorruso, bikol, bugis, dugri, armenio, kimbundu, luyia, flamenco occidental y tailandés meridional.

Figura 2: El proceso de entrada / salida del modelo.



El proceso del modelo se basa en mecanismos de ponderación capaces de transformar cifras por país en cifras por lengua, técnicas de extrapolación para complementar fuentes que ofrecen valores para un número limitado de países y mecanismos de ponderación con la distribución mundial de personas conectadas a la Internet, con el fin de producir porcentajes globales de las diversas fuentes.

Tabla 1 : Los 2 tipos de ponderaciones utilizados.

	Demo-lingüística	Usuarios de la Internet por lengua
ENTRADA	% por país ----> % por lengua	% Criterio ---->% mundial
SALIDA	Datos por país	Dado en% según criterios específicos
PRODUCCIÓN	Datos por lengua	Datos relativos a la población L1 + L2
DATOS DE PONDERACIÓN	Hablantes L1 + L2 por país	% de personas conectadas a la Internet por país
ALCANCE	Todas las fuentes por país	Indicadores de índice e interfaz.
HIPOTESIS IMPLÍCITA	Independencia de lenguas en el país	Tasa de modulación de la conexión a la Internet según criterio

El modelo se implementa en Excel en una hoja de cálculo de 7 MB con 17 pestañas correlacionadas, organizadas en torno a los 215 países considerados, las 138 lenguas procesadas y los 412 micro-indicadores recopilados. El modelo así configurado permite verificar en una fracción de segundo el impacto de cualquier hipótesis (incluido el análisis prospectivo).

1. INTRODUCCIÓN

Esta segunda versión del método para la creación de indicadores de presencia de lenguas en la Internet trae un conjunto de mejoras tangibles que impactan positivamente en la confiabilidad del método y reducen los sesgos.

Las principales mejoras se derivan de la adopción del *Ethnologue Global Dataset 24*⁴, de marzo de 2021, que no solo actualiza los datos demo-lingüísticos (la cantidad de hablantes de cada lengua en cada país) sino que también brinda los datos más fiables sobre el tema, incluso si la precisión perfecta es inaccesible, y, además, ofrece la primera fuente histórica para la distribución del número de hablantes de L2 por país, para cada lengua.

2. DIFERENCIAS CON LA PRIMERA VERSIÓN

Se han realizado muchas diferencias en el método o las fuentes en comparación con la versión 1, con el ánimo de mejorar la calidad del método y los productos.

2.1 Elección de los datos de Ethnologue como fuente demo-lingüística

La mayor parte de la fuente proporcionada por Ethnologue está en forma de matriz Excel de 11.500 líneas en el siguiente formato: “ISO639⁵, Nombre de la lengua, nombre del país, número de hablantes L1, número de hablantes L2, más una gran cantidad de parámetros asociados no utilizados para este método”.

Para obtener el formato requerido por el modelo (una matriz con todos los países considerados en columna y todas las lenguas considerados en filas), se ha implementado un conjunto de pasos cuidadosos con el apoyo de diferentes macros escritas en VBA⁶. Uno de los pasos más complejos fue fusionar todos los datos de lenguajes pertenecientes al mismo macrolenguaje. Este proceso involucró 60 macro-lenguas agrupando 434 lenguas diferentes.⁷ (ver detalles en el anexo 2).

Después de completar este paso, el proceso consistió en reducir la lista de lenguas para conservar solo las que maneja el modelo⁸, agregando cuidadosamente todos los datos restantes por país en una sola fila "RESTO".

Es importante entender que la adopción de los datos de Ethnologue implica la aceptación de los formatos de presentación, los cuales se basan en consideraciones puramente lingüísticas:

- Agrupación de macro-lenguas⁹
- Lista de países y denominación inglesa correspondiente.

La lista de países tratados por Ethnologue es mayor que la UIT¹⁰ usa para las tasas de conexión a la Internet por país: la UIT, como entidad de las Naciones Unidas, no separa, por ejemplo, Martinica de

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

⁵ El código ISO de 3 caracteres asignado a cada una de las 7486 lenguas identificadas.

⁶ Virtual basic for Applications

⁷ Por ejemplo, la macro-lengua árabe incluye 29 lenguas, como el árabe egipcio o el árabe marroquí.

⁸ En este punto 138 lenguas con más de 5 millones de hablantes de L1.

⁹ Un ejemplo significativo es el caso de la macro-lengua serbocroata, cuya definición agrupa, en orden alfabético, el bosnio, el croata, el montenegrino y el serbio. Este reagrupamiento obligado no responde en absoluto criterios geopolíticos e incluso podría considerarse controvertido desde este punto de vista. Además, como algunas fuentes separan claramente las lenguas y países involucrados, existe el riesgo de error en los resultados, incluso si la entrada de fuentes se ha transformado para tener en cuenta esta situación (el riesgo surge cuando las cifras no se deben agregar sino promediar como en el indicador de profundidad de Wikipedia).

¹⁰ La Unión Internacional de las Telecomunicaciones (<http://itu.int>), el organismo de las Naciones Unidas que proporciona estadísticas sobre telecomunicaciones, incluido el porcentaje de personas conectadas a la Internet por país.

Francia). En este caso, la regla de la UIT es la que debemos adoptar y el requisito ha sido recopilar cuidadosamente los datos de Ethnologue para los 29 países no considerados por la UIT (para la lista completa, ver el anexo 3) en una sola columna “Otros países”.

2.2 Gestión de L2 y multilingüismo

La inclusión de los últimos datos del Ethnologue en el modelo permitió, como subproducto, eliminar el mayor sesgo del método vinculado al tratamiento de L2. Por primera vez, existe una fuente confiable que complementa, para cada lengua, el número de hablantes L1 por país con el número de hablantes L2 por país. En la versión 2017, las cifras para personas L2 conectadas se calcularon a partir del total de hablantes L2 en el mundo aplicando la misma tasa de conectividad obtenida por el modelo para parlantes L1. Un sesgo significativo se debe al hecho de que para ciertas lenguas importantes (como el francés y el inglés) una alta proporción de hablantes L2 pertenecen a países en desarrollo donde la tasa de conexión promedio es mucho más baja que la obtenida en promedio para hablantes L1.

Otra consecuencia positiva del uso de los datos de Ethnologue es la posibilidad de obtener una “cifra oficial” del multilingüismo. La ratio mundial $(L1 + L2) / L1$ se estableció en la edición de 2017 proyectando los datos disponibles para los países tratados: salió alrededor de 1,25. Ahora el número lo proporcionan indirectamente los datos de Ethnologue y su valor es 1,43.

Los datos del Ethnologue son los siguientes:

- ✓ Población mundial (total mundial de hablantes L1): 7.231.699.136
- ✓ Total mundial de hablantes L1 + L2: 10.361.716.756
- ✓ La "tasa de multilingüismo" es, por tanto, $10.361.716.756 / 7.231.699.136 = 1,4328$
(en otras palabras, el 43% de la población mundial es al menos bilingüe).

Esta cifra de 43% es mucho mejor que el 25% utilizado en la primera versión y no es un elemento anecdótico del modelo sino uno de los elementos clave. Como muestra el primer estudio, el sesgo más común y crítico en las cifras de lenguas ofertadas es que no consideran correctamente a los hablantes L2 (un problema que se expresa plenamente en la Internet donde la mayoría de los Internautas utilizan su segunda lengua y donde muchos sitios web son multilingües¹¹). No prestar atención a esto conduce a errores enormes, a menudo ocultos en "el resto de las lenguas", ya que los porcentajes mundiales que deben calcularse para una población de 10 mil millones (hablantes de L1 + L2) se calculan sobre un total de 7 mil millones (la población mundial).

En esta segunda edición del método, el principio de medir todo en términos de la población L1 + L2 (en lugar de la población mundial) se adoptó por completo para garantizar la precisión de los resultados. Por ello (y también por otras mejoras), la comparación entre los resultados de 2017 y los de 2021 debe realizarse con cautela. De hecho, todos los macro-indicadores, *potencia*, pero también *capacidad* y *gradiente* siguen ahora esta regla de ser calculados sobre la población L1 + L2 en lugar de la población L1 y, por lo tanto, aparecerán con valores más bajos en comparación con la versión 2017.

¹¹ De hecho, los 6 indicadores tratados por el estudio son por naturaleza multilingües: los usuarios de la Internet visitan sitios y generan tráfico en las diferentes lenguas que dominan, a menudo los sitios web son multilingües, las interfaces son multilingües, los servicios de traducción cubren diferentes lenguas.

2.3 Fuente para personas conectadas a la Internet

Hasta 2017, la UIT proporcionó una actualización anual de sus datos¹² sobre el porcentaje de personas que utilizan la Internet por país, incluidas sus propias estimaciones cuando no existía una fuente oficial para algunos países. Estos datos, que son un elemento central del método, se consideraron entre los más fiables. Desafortunadamente, después de 2017, la UIT decidió dejar de proporcionar sus propias estimaciones, dejando a muchos países (casi todos ellos en vías de desarrollo¹³) con las antiguas cifras de 2017.

Esto supuso un grave problema para este estudio y, tras algunas iteraciones, llevó a la decisión de derogar, en este caso, un principio fundamental en este tipo de trabajo estadístico: el de no modificar nunca los datos de las fuentes.

El Banco Mundial proporciona sus propias cifras¹⁴ para el mismo indicador, que son claramente tomados de la UIT, pero en varios casos superan la limitación actual y proponen nuevas cifras donde la UIT dejó los datos de 2017. Esto es un avance, pero muchos países aún permanecen fuera de la actualización de 2017, lo que tendría un impacto negativo en las lenguas habladas en estos países e impide percibir progresos potenciales.

Finalmente, se decidió utilizar los datos del Banco Mundial cuando complementen a los de la UIT y, para los muchos casos restantes no actualizados, realizar, para cada país en cuestión, una búsqueda en la Internet de datos fiables y proporcionar estimaciones basadas, a menos que se indique lo contrario, en la progresión lineal aproximada de los datos de años anteriores.

Un caso sigue siendo un problema a pesar de todo: India tiene ahora una cifra oficial del 20,1% en 2021 mientras que la estimación de la UIT en 2017 era del 32% y muchas fuentes informan de un boom de la Internet en este país con cifras en torno al 50%¹⁵! Al no lograr obtener una respuesta de la fuente oficial o de los colegas indios consultados, se decidió, debido a la importancia primordial de la India en el contexto del estudio¹⁶, derogar excepcionalmente un principio aún más fuerte: el de no modificar las fuentes oficiales. El supuesto es que la cifra proporcionada por el Ministerio de Estadísticas e Implementación de Programas de la India se refiere únicamente a las conexiones fijas y no incluye las conexiones móviles a la Internet. Con base en este supuesto, se estableció la cifra conservadora del 40%. Se debe tener en cuenta que la sensibilidad de esta cifra a los resultados no es marginal. A continuación, se muestran los diferentes resultados finales para hindi y bengalí según el valor del porcentaje de personas conectadas a la Internet en la India.

Tabla 2: Sensibilidad de las cifras de la India sobre el porcentaje de personas conectadas a la Internet

% personas conectadas en India	20,08%	30%	40%	50%
Potencia del hindi (rango)	2,42% (10)	2,91% (8)	3,38% (5)	3,81% (4)
Potencia del bengalí (rango)	0,75% (17)	0,82% (15)	0,88% (15)	0,95% (14)

¹² <https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2021/PercentIndividualsUsingInternet.xlsx>

¹³ Solo 80 países proporcionaron cifras oficiales en 2019.

¹⁴ Fuente: <https://data.worldbank.org/indicador/IT.NET.USER.ZS>

¹⁵ Ver por ejemplo <https://www.statista.com/estadísticas/255146/número-de-usuarios-de-internet-en-india/> o entonces https://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users

¹⁶ Con 34 lenguas que forman parte de la lista de las lenguas consideradas, incluidos lenguas myores como el hindi y el bengalí.

2.4 Gestión de fuentes para micro-indicadores

Todo el proceso de gestión de las fuentes de los micro-indicadores es la tarea más pesada y difícil del proyecto, con un alto consumo de recursos humanos puesto que son necesarios muchos pasos:

1. Para cada indicador, buscar fuentes en la Internet.
2. Seleccionar las fuentes según su fiabilidad y su aplicabilidad al proceso previsto¹⁷.
3. Recopilar las fuentes seleccionadas en un formato válido para el proceso automático.
4. Introducir las fuentes en el modelo.
5. Evaluar los sesgos de las fuentes.

En el anexo 1, puede consultarse la lista completa de fuentes para cada indicador.

Para realizar el paso 4, los datos deben ser transformados al formato Excel, con los nombres de países y lenguas correspondientes a los del modelo y en el mismo orden secuencial.

Para el paso 3, todas las fuentes se recopilan de una URL específica (consulte el Anexo 1 para obtener la lista completa de URL) y la mayoría de las fuentes se obtienen en formato HTML. Algunas fuentes están en formato PDF y un subconjunto limitado (principalmente el de la UIT y el Banco Mundial) está en formato Excel, con el cual trabaja el modelo. El proceso de conversión de PDF a formatos compatibles puede ser relativamente sencillo en la mayoría de los casos, pero en algunos casos existe una incompatibilidad y se necesitan gestiones y programaciones adicionales.

El proceso de transformación de HTML a Excel a menudo puede convertirse en una verdadera pesadilla que requiere mucha imaginación y trucos, incluso, en algunos casos, la necesidad de ir a recuperar los datos dentro de la fuente HTML e intentar desde allí construir una tabla que se adapte al modelo, después de limpiar el código HTML.

En un número cada vez mayor de casos, la fuente ofrece acceso geográfico a los datos (mapas interactivos) los cuales, excepto cuando el número de países o lenguas es limitado y la copia manual no es demasiado pesada, hace que el procesamiento automático sea imposible o requiera la subcontratación para una persona de confianza para un trabajo de recolección manual que es tedioso y requiere una gran concentración y disciplina para evitar errores. Por tanto, se subcontrató la recopilación de datos de tráfico que incluían cientos de micro-indicadores.

Se debe dar crédito a las instituciones (en general, organizaciones internacionales u ONG) que proporcionan los datos en un formato utilizable por computadora (Wikimedia proporciona, por

¹⁷ Puede ocurrir que una fuente sea confiable pero expresada en un formato que hace imposible automatizar su colecta.

ejemplo, en su versión en inglés, tablas HTML que se transforman directamente en formato compatible con Excel, sin pérdida de estructura).

Obtener una copia de la fuente en Excel o en un formato compatible (generalmente una matriz de nombres de países o lenguas con valores o porcentajes asociados) no es el final del proceso. Con 250 países y cientos de lenguas para procesar y en lugar del código ISO, inequívoco, el uso común de nombres literales que pueden estar en diferentes lenguas y en grafías no estándar, la integración de los datos en el modelo no puede hacerse a mano. Se han escrito dos macros para este proceso que en ambos casos requieren un ajuste recursivo¹⁸ para integrar las diferentes grafías. Las salidas de estos programas son archivos Excel que se pueden usar directamente para copiar las fuentes en su totalidad, o línea por línea, en el archivo Excel del modelo. Además del enorme ahorro de tiempo de este método, garantiza obtener los datos sin errores.

También se debe tener en cuenta que el manejo de las macro-lenguas ha hecho que este proceso sea aún más complejo, porque la agrupación de lenguas debe realizarse en los datos de origen antes de procesarlos por la macro. Para citar algunos ejemplos, las frecuentes apariciones de árabe egipcio o marroquí en las fuentes se acumularon en la macro-lengua árabe y las de serbio, bosnio, croata y montenegrino se fusionaron en serbo-croata (el número de casos similares es bastante elevado). Para el procesamiento manual de las salidas del programa señalando lenguas desconocidas, se utilizó como soporte la página Ethnologue que describe cada código de lengua¹⁹.

2.4.1 INDEXES

La fecha límite llegó demasiado pronto durante la producción de la versión 2017 y este indicador resultó insuficiente con una sola fuente que proporciona 5 micro-indicadores. En esta ocasión, se ha prestado la atención necesaria y se ha realizado una recopilación de datos casi exhaustiva para este indicador. Se han incluido una amplia variedad de parámetros que caracterizan el progreso de los países en la sociedad de la información, con ahora 25 micro-indicadores, desde la estabilidad eléctrica hasta la inteligencia artificial, la gobernanza y muchos otros parámetros (ver el anexo 1 para la serie completa).

2.4.2 CONTENIDOS

Como se explicó anteriormente, las fuentes de datos sobre lenguas en la Internet son extremadamente escasas y, por lo tanto, este indicador depende en gran medida de las estadísticas excepcionales de Wikimedia. Sin embargo, en el capítulo sobre sesgos del documento se establece que la presencia de lenguas en Wikimedia no es proporcional a su presencia en la Web y por tanto no puede considerarse como un indicador fiable de la distribución lingüística de los contenidos de la Internet.

Se ha introducido en el modelo una forma de equilibrar los resultados de las estadísticas de Wikimedia, pero el doloroso diagnóstico es que el contenido sigue siendo el indicador más

¹⁸ El proceso recursivo finaliza cuando el proceso de origen ya no produce una ortografía desconocida.

¹⁹ <https://www.ethnologue.com/language/srp>

débil de este método, aunque es un elemento muy sensible (los cambios relativamente marginales en los valores de este indicador pueden provocar un gran impacto en los valores de los macro-indicadores). Si bien uno de los principales objetivos del proyecto es, por supuesto, conocer la distribución lingüística del contenido de Internet, existe una dificultad frustrante para ponderar correctamente los contenidos de la Internet; en ese contexto el macro-indicador holístico *potencia* queda siendo la mejor aproximación a los contenidos²⁰.

Para intentar controlar mejor la excesiva influencia de las cifras de Wikimedia sobre este indicador, se tomaron dos decisiones. La primera afecta exclusivamente a la enciclopedia Wikipedia: en lugar de tener un indicador para cada una de las cifras disponibles (número de artículos, editoriales activas, ediciones y profundidad²¹) se ha creado una fórmula para definir un único micro-indicador que integra debidamente cada factor:

$$W(Li) = \text{Artículos}(i) \times \text{Ediciones}(i) \times \text{Editores}(i) \times \text{Profundidad}(i) / L1 + L2(i)^2$$

Esta fórmula expresa con mayor precisión la actividad general de Wikipedia por lengua, lo que reduce el peso de lenguas cuando se utilizan *bots* (programas que simulan el comportamiento de humanos) para crear artículos a partir de la copia y traducción de artículos existentes en otra lengua, los artículos obtenidos quedándose a menudo sin más actualización²².

La siguiente tabla muestra cómo esta fórmula refleja mejor la realidad. La última columna (presencia), que es la relación entre el número de artículos y la población L1 + L2 (número de artículos por hablante) es una clara demostración de por qué la presencia de lenguas en Wikipedia no es un buen indicador de la presencia global de lenguas en la Internet...

²⁰Como muestra la primera edición, el loable esfuerzo de W3Techs para obtener cifras de contenidos se caracteriza por sesgos muy significativos y en muchos niveles (el más fuerte, pero no el único, es la falta de consideración del multilingüismo y el hecho de que los sitios web multilingües que incluyen el inglés son probablemente contados solo en inglés). Así, esta fuente proyecta valores para los contenidos en inglés que son extremadamente exagerados (por encima del 50% mientras que la realidad probablemente esté hoy por debajo del 25%). La falta de múltiples fuentes mantiene así el mito en los medios de que más de la mitad de los sitios web están en inglés. Este fue el caso entre 2007 y 2009 (ver [3]), pero desde el crecimiento exponencial del chino, hindi, árabe, turco, bengalí, vietnamita, urdu, persa y marathi, por nombrar los que se encuentran en los primeros 20 lugares y que en conjunto representan casi el 28% de los contenidos, ha cambiado radicalmente la situación y el inglés representa solo una cuarta parte del contenido. Entre 2000 y 2007, el mito persistente era que el inglés ocupaba el 80% de la Web y esta desinformación finalmente desapareció después de 2009 con la publicación por parte de la UNESCO de informes (ver [3] y [4]) que establecieron una presencia del inglés en torno al 50%. ¿Cómo pudo el inglés haberse mantenido estable al 50% de los contenidos durante 14 años con todos los cambios que ha experimentado la Internet y una cantidad de angloparlantes conectados (L1 + L2) que ahora se reduce al 13% de todas las personas conectadas?

²¹ Citado de Wikimedia: *Profundidad*, que se define como $[\text{Modificaciones} / \text{Artículos}] \times [\text{No artículos} / \text{Artículos}] \times [1 - \text{Stub-ratio}]$, es un indicador aproximado de la calidad de una Wikipedia, que muestra la frecuencia con la que se actualizan sus artículos. No se refiere a la calidad académica.

²² Sin esta fórmula, el cebuano, con un gran número de artículos, pero muy poca profundidad, aparece con la mayor puntuación de *gradiente*.

Tenga en cuenta que el valor de profundidad para el vietnamita no ha sido informado y hemos puesto un valor de 1 para evitar una fórmula nula²³.

Tabla 3: Factores de Wikipedia y la fórmula

Lengua	Artículos	Modificaciones	Usuarios activos	Profundidad	FÓRMULA	PRESENCIA
Inglés	6332139	1027716498	125399	1073	481775	0,47
Cebuano	5853095	32075254	186	2	275	36,71
Sueco	3050759	49330695	2148	12	22759	23,37
Alemán	2593827	212207089	18119	93	50897	1,92
Francés	2342875	183969129	18054	242	26424	0,88
Holandés	2060512	59302602	3933	17	13742	8,45
Ruso	1736736	115035192	10425	137	4286	0,67
Italiano	1703284	121418801	8085	172	62435	2,51
Español	1698331	136390848	15694	210	2590	0,31
Polaco	1480982	63723938	4235	32	7742	3,64
Japonés	1277204	84188217	15173	85	8683	1.01
Vietnamita	1266628	65110373	2476	1	35	1,65
Chino	1208732	66159632	8940	202	62	0,08
Árabe	1123561	54279052	5189	227	536	0,31
Ucranio	1100281	32831286	2773	53	4823	3.32
Portugués	1067241	61371751	9508	176	1651	0,41

En el capítulo que trata sobre el sesgo, se presenta un análisis en profundidad de las estadísticas de Wikimedia.

La segunda decisión tomada para equilibrar la influencia de Wikimedia en el indicador de *contenidos* es un sistema de ponderación implementado que le da más importancia al T-Index de Translated²⁴ que a toda la colección de indicadores de Wikimedia. Jugar con diferentes configuraciones de factores de ponderación ha demostrado la alta sensibilidad de este indicador, principalmente debido al bajo número de fuentes y al hecho de que algunas lenguas tienen una presencia desproporcionada en relación a su número de hablantes.

La configuración de ponderación finalmente implementada es la siguiente:

Tabla 4: Ponderación de los indicadores de contenido

OBJETO	PESO
Amazon EE.UU.: cantidad de libros 2017 ²⁵	0,5
Fórmula de Wikipedia	1

²³ El valor de profundidad bajo refleja el hecho de que el 67% de las publicaciones son creadas por bots, no humanos (fuente: https://www.wikiwand.com/en/Vietnamese_Wikipedia).

²⁴ Este índice, accesible en <https://translated.com/les-langues-qui-comptent>, es un intento de medir el potencial de los idiomas en el comercio electrónico a partir del número de usuarios de la Internet por idioma, multiplicado por el gasto estimado en línea. Utilizó cifras del Banco Mundial y la UIT y propuso una proyección 2021 que es la cifra utilizada para el modelo. Además de los datos de Wikimedia, es una de las pocas fuentes serias de idiomas disponibles en la Internet.

²⁵ La falta de datos accesibles equivalentes para 2021 y la situación con Wikimedia llevaron a la decisión de mantener este micro-indicador aunque no está actualizado.

Número de WikiBooks por lengua	0,5
Artículos de WikiQuote por lengua	0,1
Número de artículos de WikiSource por lengua	0,1
Número de artículos de Wikiversity por lengua	0,1
Número de artículos de Wikcionario por lengua	0,1
Número de artículos de WikiNews por lengua	0,1
Número de artículos de WikiVoyages por lengua	0,1
Proyección del T-index para el comercio electrónico en 2021	3

2.4.3 TRÁFICO

El trabajo para el indicador de *tráfico* también fue muy pesado con muchos ensayos y errores. En 2017, se determinó que los datos de Alexa (porcentaje de tráfico por país a una serie de sitios) estaban extremadamente sesgados en contra de los países asiáticos (especialmente India y China) y Brasil y algo sesgados a favor del francés y el inglés. Cuatro años después, la recopilación de datos de Alexa mostró situaciones extrañas, algunas métricas no reportaban el tráfico en el país donde se creó el sitio²⁶ y la impresión de una tendencia a subestimar los tráficos de los países europeos, por otro lado, India aparece bastante alto en todos los sitios lo que no ocurre con China.

Un estudio que comparó los datos de tráfico con los datos de suscripción de cinco redes sociales importantes confirmó las impresiones intuitivas. En resumen, el tráfico procedente de Brasil parece muy subestimado en comparación con el nivel de suscripción, lo mismo para Alemania, España, Francia, Italia y Reino Unido; por otro lado, India, Japón y Corea parecen estar muy sobreestimados. Consultar el capítulo sobre sesgos para obtener más detalles.

Ante estos resultados poco alentadores, se decidió buscar una herramienta de medición alternativa. SimilarWeb parecía ser la mejor alternativa y la prueba se planeó antes de comprar una suscripción. Desafortunadamente, fue imposible acceder a los datos de tráfico por país y, a pesar de los numerosos intentos de comunicarse a través de diferentes canales, incluido el chat interactivo de la empresa, nunca se obtuvo la mínima respuesta.

Ante esta situación de estancamiento, se probó otro proveedor, Semrush.com, y se recopilieron cifras por país para los mismos sitios web. Semrush, a diferencia de Alexa, proporciona, para cada sitio medido, resultados para todos los países, lo que fue una perspectiva interesante, eliminando la necesidad de extrapolación. Sin embargo, sucede que en algunos casos el total es menor al 100% (lo cual no es un problema) y otras veces supera el 100% (lo cual es un problema). Las cifras se han normalizado para que sean 100% precisas mediante una regla de prorrateo antes de ingresar al modelo.

²⁶ Por ejemplo, theses.fr mostró tráfico cero en Francia, al igual que spip.net, un CMS utilizado principalmente en Francia.

Después de ejecutar el modelo, transformando los datos nacionales en datos lingüísticos, los resultados no fueron convincentes: el valor del chino era demasiado bajo, lo mismo para el hindi y el árabe y para las “lenguas restantes”.

Las diferencias extremas entre los resultados obtenidos a partir de los datos de Alexa y Semrush son una señal de alerta sobre la confiabilidad de estas herramientas y una preocupación por los planes futuros para expandir el número de sitios web estudiados y permitir la diferenciación temática para algunas lenguas.

2.4.4 INTERFACES

La lista de lenguas aceptadas en las interfaces de aplicaciones importantes o como posibles destinos para los servicios de traducción en la web no plantea problema particular. La lista de aplicaciones seleccionadas se puede ver en el Anexo 1. Tenga en cuenta que, para reducir la importancia de los datos de Wikimedia en el modelo, se decidió eliminar las fuentes de Wikimedia de este indicador.

2.4.5 USOS

Tampoco hay dificultades particulares para este indicador, aparte la de encontrar datos gratuitos (principalmente número de suscriptores por país) de las principales redes sociales. Finalmente, se logró incluir datos para las siguientes aplicaciones: Facebook, Instagram, LinkedIn, Messenger, Pinterest, Reddit y Twitter. Además, se han incluido algunas fuentes ajenas a las redes sociales, como el número de descargas de OpenOffice por país (ver la lista completa en el Anexo 1).

2.5 Resumen de indicadores

La siguiente tabla resume la descripción de cada uno de los indicadores y cómo se construye a partir de micro-indicadores.

Tabla 5: Descripción de indicadores

INDICADOR		TÉCNICO	FIABILIDAD / SESGO
A: INTERNAUTAS	Indicador único basado en cifras de la UIT y el Banco Mundial para el % de personas conectadas por país, extrapolado cuando faltan cifras.	Ponderación país -> lengua sin extrapolación	Alta fiabilidad Sesgo muy marginal aunque está aumentando debido a la falta de actualización en muchos países.
B: USOS	Incluye 14 micro-indicadores con datos de 2021: - % fijo + móvil por país - % de banda ancha por país - Descarga acumulativa de OpenOffice - Facebook, Instagram, LinkedIn, Messenger, Netflix, Pinterest Twitter, YouTube, % de suscriptores por país	Ponderación país -> lengua extrapolado en proporción de velocidad de conectividad Promedio de micro-indicadores	Fuerte confiabilidad. Sesgo bajo.

C: TRÁFICO	Alexa midió el tráfico por país a una selección de 338 sitios web.	ponderación país -> lengua extrapolado proporcionalmente Promedio truncado al 20%	Fiabilidad relativamente buena Pero los fuertes sesgos europeos negativos de Alexa se confirman mediante comparaciones de tráfico y número de suscriptores por país.
D: INDEX	Incluye 25 indexes de varias fuentes que miden parámetros como: - E. gobierno - Acceso universal - E. participación - Infraestructura general (Ver Anexo 1 para la lista completa)	ponderación país -> lengua extrapolado por el método de los cuartiles. Luego, transformación en datos globales por ponderación porcentual con la UIT Promedio de micro-indicadores	Buena fiabilidad y sesgo marginal (datos subjetivos cuantificados por un organismo competente).
E: CONTENIDOS	Incluye 13 micro-indicadores con ponderación asociada. T-Index de Translated, una medida del potencial de comercio electrónico de una lista de lenguas (2021) - Número de libros en Amazon (2017) - 11 micro-indicadores de lengua de Wikimedia: artículos, usuarios o editores; todos los indicadores de Wikimedia se sintetizan con una fórmula.	Uso directo de dígitos ponderados por lengua para equilibrar la importancia de Wikimedia. Fusión de 4 indicadores de Wikipedia con una fórmula. Promedio truncado al 20% del micro-indicador	Muy buena fiabilidad para Wikimedia y Amazon. Pero bastante sesgado debido a la escasa presencia de algunos de las principales lenguas asiáticas. Debería aumentarse el número de micro-indicadores para dar más fuerza al promedio.
F: INTERFAZ (e lenguas de traducción)	Incluye 23 micro-indicadores binarios	% de presencia en los 23 micro-indicadores. % de ponderación mundial con cifras de la UIT.	Perfecto.

3. RESULTADOS

Las siguientes tablas presentan los resultados para las lenguas más altas en cada macro-indicador, excluyendo la productividad²⁷. La siguiente tabla muestra todos los resultados resumidos para las 15 lenguas más "potentes" de la Internet. Los resultados están expresados en porcentajes sobre la base de la población L1 + L2. Conn.M significa Porcentaje Mundial de Personas Conectadas, Pop.M significa Población Mundial y L. Conn. Porcentaje de parlantes conectados.

Tabla 6 : Indicadores para los 15 lenguas principales en términos de potencia

	Conn.M	Pop.M	TRÁFICO	L. Con.	USOS	CONT.	INTERF.	INDEX	POTENCIA	Capac.	Grad.
Inglés	15,30%	13,01%	37,44%	64,33%	27,92%	38,61%	21,73%	17,87%	26,48%	2,04	1,73
Chino	17,65%	14,72%	7,79%	65,59%	5,47%	8,18%	25,07%	19,38%	13,92%	0,95	0,79
Español	7,00%	5,24%	10,72%	73,08%	11,74%	5,42%	9,94%	7,59%	8,73%	1,67	1,25

²⁷Este indicador se volverá a evaluar en la sección Corrección de sesgos. El indicador de *potencia*, que integra todos los elementos, probablemente sería, en esta etapa, una mejor aproximación a la distribución de *contenidos* por lengua, dato que sigue siendo muy difícil de obtener de forma fiable hasta el día de hoy.

Francés	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
Portugués	3,05%	2,49%	1,42%	67,16%	5,53%	3,30%	3,85%	2,92%	3,35%	1,35	1,10
Ruso	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
Alemán	2,09%	1,30%	1,32%	87,65%	1,95%	5,84%	2,97%	2,98%	2,86%	2,19	1,37
Japonés	2,07%	1,22%	1,98%	92,62%	1,76%	3,55%	2,77%	3,01%	2,52%	2,07	1,22
Malayo	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
Bengalí	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%		

La fila RESTO representa el conjunto completo de todas las lenguas del mundo, excepto las 15 lenguas enumerados en la tabla. Debe quedar claro que la clasificación en términos de *potencia* favorece a las lenguas que tienen el mayor número de hablantes. Los macro-indicadores de *capacidad* y *gradiente* proporcionan resultados independientes del número de hablantes.

Recordatorio:

Potencia²⁸ se definió como el promedio de los 5 indicadores.

Capacidad²⁹ es el valor de la *potencia* dividido por el % de parlantes L1 + L2

Gradiente³⁰ es el valor de la *potencia* dividido por el % de parlantes L1 + L2 conectados

La siguiente tabla muestra las lenguas más conectadas.

Tabla 7 : Lenguas ordenados por porcentaje de personas conectadas

CLASIFICACIÓN INTERNAUTAS	Internautas	Capacidad	Gradiente
Danés	97,82%	2,19	1,22
Sueco	93,49%	2,61	1,53
Japonés	92,62%	2,07	1,22
Holandés	92,02%	2,26	1,34
Suizo alemán	91,56%	1,21	0,72
Flamenco occidental	90,43%	1,12	0,68
Finlandés	89,67%	3,42	2,09
Bávaro	87,68%	0,97	0,61
Alemán	87,65%	2,19	1,37
Hebreo	85,46%	5,24	3,35
Eslovaco	82,47%	1,30	0,86
Bielorruso	82,27%	1,00	0,66

²⁸Se ha utilizado la palabra potencia en lugar de ponderación para evitar confusiones con el importante uso transversal de ponderación en el método. Representa la presencia absoluta de un idioma en Internet integrando todos los factores.

²⁹La capacidad es la presencia relativa de un idioma en Internet, independientemente de su número de hablantes; indica el dinamismo de un idioma en Internet.

³⁰El gradiente indica el dinamismo de los altavoces conectados; Se eligió el término gradiente, que expresa una derivada y, por lo tanto, una tendencia, porque un gradiente fuerte es una promesa de aumento de la capacidad.

Checo	81,37%	1,70	1,14
Polaco	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
Serbocroata	77,78%	3,14	2,21
Griego	77,71%	1,75	1,23
Ruso	77,20%	1,25	0,88
Kazajo	76,98%	0,90	0,64
Rumano	75,66%	1,18	0,86
Italiano	75,65%	2,09	1,51
Albanés	75,48%	1,12	0,81
Azerbaiyano	74,76%	0,94	0,69
Napolitano-calabrés	74,39%	0,84	0,62
Español	73,08%	1,67	1,25
Kurdo	73,02%	0,89	0,67
Búlgaro	70,34%	1,18	0,92
Armenio	69,86%	1,41	1,11
Vietnamita	69,04%	1,07	0,85
Guaraní	68,83%	0,64	0,51
Portugués	67,16%	1,35	1,10

La siguiente tabla está ordenada por *capacidad*.

Tabla 8 : Lenguas ordenados por capacidad

CLASIFICACIÓN POR CAPACIDAD	usuarios de Internet	Capacidad	Gradiente
Hebreo	85,46%	5,24	3,35
Finlandés	89,67%	3,42	2,09
Serbocroata	77,78%	3,14	2,21
Sueco	93,49%	2,61	1,53
Holandés	92,02%	2,26	1,34
Alemán	87,65%	2,19	1,37
Danés	97,82%	2,19	1,22
Italiano	75,65%	2,09	1,51
Japonés	92,62%	2,07	1,22
Inglés	64,33%	2,04	1,73
Polaco	81,17%	1,88	1,26
Húngaro	79,92%	1,79	1,22
Griego	77,71%	1,75	1,23
Checo	81,37%	1,70	1,14
Español	73,08%	1,67	1,25
Turco	77,98%	1,46	1,02
Francés	63,67%	1,44	1,24
Armenio	69,86%	1,41	1,11
Portugués	67,16%	1,35	1,10
Eslovaco	82,47%	1,30	0,86
Ruso	77,20%	1,25	0,88

Y finalmente, la última tabla, ordenada por *gradiente*, destaca el dinamismo de las personas conectadas. La presencia en tercera posición del malgache³¹ es consecuencia del dinamismo de sus hablantes en algunos indicadores de Wikimedia.

Tabla 9 : Lenguas ordenados por gradiente

CLASIFICACIÓN POR GRADIENTE	Internautas	Capacidad	Gradiente
Hebreo	85,46%	5.24	3.35
Serbocroata	77,78%	3,14	2.21
Malgache	9,79%	0,40	2.21
Finlandés	89,67%	3,42	2,09
Inglés	64,33%	2,04	1,73
Sueco	93,49%	2,61	1,53
Italiano	75,65%	2,09	1,51
Alemán	87,65%	2.19	1,37
Holandés	92,02%	2,26	1,34
Polaco	81,17%	1,88	1,26
Español	73,08%	1,67	1,25
Francés	63,67%	1,44	1,24
Griego	77,71%	1,75	1,23
Danés	97,82%	2.19	1,22
Húngaro	79,92%	1,79	1,22
Japonés	92,62%	2,07	1,22
Checo	81,37%	1,70	1,14
Armenio	69,86%	1,41	1,11
Portugués	67,16%	1,35	1,10

Más allá del previsible hecho de que las lenguas nacionales de países reconocidos por sus políticas proactivas a favor de la sociedad de la información ocupen los primeros lugares, es destacable que varias lenguas se ubiquen por encima del inglés, a pesar de su ventaja estratégica en la Internet (lengua de elección para contenidos multilingües y el hecho que muchos creen que es la lengua franca de la Internet).

Estos resultados deben tomarse con ciertas reservas debido a los sesgos mencionados en el documento, en particular las dificultades con el indicador de *contenidos*, cuyas variaciones pueden tener un impacto considerable en estos macro-indicadores³².

4. ANÁLISIS DE LOS RESULTADOS

³¹ Una clasificación de este tipo para el malgache, una lengua con menos del 10% de hablantes conectados y una capacidad muy baja, puede causar una sorpresa legítima: es el resultado de un "accidente matemático" debido a una presencia extremadamente desproporcionada en la lengua. de los micro-indicadores de contenido y es un síntoma de la debilidad de este indicador que se comenta a continuación.

³² Antes de la introducción de la fórmula de Wikipedia y la ponderación de Wikimedia, aparecía en la primera fila el cebuano, el segundo idioma en términos de número de artículos de Wikipedia, cercano al inglés, con un número de artículos dos órdenes de magnitud mayor que su número de hablantes.

Aunque las comparaciones con los resultados de 2017 deben hacerse con cautela debido al tamaño y la naturaleza de los cambios (en particular, la opción de expresar porcentajes en relación con la población mundial total L1 + L2), se pueden destacar ciertos fenómenos.

El esperado crecimiento del hindi que pronto podría competir con el francés por el 4º lugar y la aparición del turco en la lista de las lenguas más potentes. Como era de esperar también, las diferencias entre las lenguas que siguen al francés y al hindi son demasiado pequeñas para considerar que los resultados están más allá del intervalo de confianza: portugués, ruso, árabe y alemán. Sin embargo, la demografía podría en un futuro próximo separar las respectivas posiciones al ritmo de la reducción de la brecha digital.

En cuanto a los macro-indicadores independientes del número de hablantes, la aparición del serbocroata debe tomarse con precaución por el riesgo de errores en la gestión de fuentes derivados de la decisión de adoptar la clasificación Ethnologue para macro-lenguas. Y claramente, el indicador de *contenidos*, y su fuerte dependencia de las estadísticas de Wikimedia, a pesar del esfuerzo realizado para contrarrestarlo, favorece claramente a las lenguas cuyos hablantes han invertido en la presencia de Wikimedia. Consulte la tabla a continuación para estas lenguas, primero ordenadas por la proporción 1000 x Número de artículos / hablantes de L1 + L2 y luego ordenados por el resultado de la fórmula implementada (factor).

Tabla 10: Presencia de lenguas en Wikipedia

Lengua	Artículos	Ediciones	Usuarios activos	Profundidad	FACTOR	% FACTOR / L1 + L2	% FACTOR / CONN	ART / L1 + L2
Sueco	3050759	49330695	2148	12	22759	1,74	1,86	233,68
Finlandés	512026	19813368	1752	40	21354	3,70	4,13	88,74
Holandés	2060512	59302602	3933	17	13742	0,56	0,61	84,51
Serbocroata	1514114	78699318	1959	92	53779	2,69	3,46	75,77
Bielorruso	281379	6093511	384	61	2620	0,67	0,81	71,87
Danés	267641	10777444	767	64	4486	0,80	0,82	47,64
Húngaro	489514	23958462	1561	59	6871	0,55	0,69	39,04
Polaco	1480982	63723938	4235	32	7742	0,19	0,23	36,44
Checo	484445	20095461	2242	46	5593	0,42	0,51	36,16
Ucranio	1100281	32831286	2773	53	4823	0,15	0,23	33,16
Búlgaro	273163	11023721	789	27	942	0,11	0,16	33,10
Hebreo	298053	31660591	3335	258	92147	9,82	11,49	31,75
Italiano	1703284	121418801	8085	172	62435	0,92	1,22	25,10
Alemán	2593827	212207089	18119	93	50897	0,38	0,43	19,21
Japonés	1277204	84188217	15173	85	8683	0,07	0,07	10,11
Persa	816984	32472834	5416	172	3534	0,04	0,07	9,77
Francés	2342875	183969129	18054	242	26424	0,10	0,16	8,78
Inglés	6332139	1027716498	125399	1073	481775	0,36	0,56	4,70

La siguiente tabla muestra claramente por qué algunas lenguas, como el hebreo, el finlandés y el serbocroata, lograron una ventaja en los resultados finales.

Tabla 10: Presencia en Wikipedia ordenada por valores de fórmula

Lengua	FACTOR	% FACTOR / L1 + L2	% FACTOR / CONN
Hebreo	92147	9,82	11,49

Finlandés	21354	3,70	4,13
Serbocroata	53779	2,69	3,46
Sueco	22759	1,74	1,86
Italiano	62435	0,92	1,22
Danés	4486	0,80	0,82
Bielorruso	2620	0,67	0,81
Húngaro	6871	0,55	0,69
Holandés	13742	0,56	0,61
Inglés	481775	0,36	0,56
Checo	5593	0,42	0,51
Alemán	50897	0,38	0,43
Polaco	7742	0,19	0,23
Ucranio	4823	0,15	0,23
Búlgaro	942	0,11	0,16
Francés	26424	0,10	0,16
Japonés	8683	0,07	0,07
Persa	3534	0,04	0,07

Estas consideraciones conducen naturalmente a la discusión sobre los sesgos.

5. ANÁLISIS DE SESGOS

Hay tres categorías principales de sesgos que pueden afectar los resultados:

- Sesgos específicos del método
- Sesgos de selección de fuentes
- Sesgos de fuentes

5.1 Sesgos específicos del método

Uno de los principales sesgos propios del método, que consiste en dar el mismo valor porcentual de hablantes L1 conectados a la Internet para hablantes L2, se eliminó con el cambio a datos de Ethnologue, gracias a la distribución de hablantes L2 por país. Este sesgo significativo afectaba particularmente a las lenguas con una gran población de L2 en países con una baja tasa de conectividad (francés e inglés). Este es un avance fundamental para la confianza en las cifras producidas por parte del modelo establecido.

El segundo sesgo del método es considerar que dentro de un país determinado todos los hablantes tienen el mismo porcentaje de conectividad (es decir, el porcentaje nacional de personas conectadas a la Internet se aplica al igual a todos los hablantes cuando es claramente un promedio). Este sesgo prohíbe distinguir entre hablantes de diferentes lenguas dentro de un país con el método (por ejemplo, los hablantes de catalán en España reciben el mismo porcentaje de conectividad que los hablantes de español y no se puede analizar ninguna ventaja de diferenciación (es lo mismo con el criollo de Martinique en Francia o con las muchas lenguas de la India). Se entiende intuitivamente que esta hipótesis no es cierta en muchos casos (la brecha digital nacional a menudo está vinculada a consideraciones lingüísticas) y que el

impacto de este sesgo es tanto más fuerte cuanto más pequeña es la población considerada. Se espera un efecto marginal si el método se aplica a una población de hablantes de más de 5 millones (aunque en el caso de la India esto puede no ser tan obvio). El próximo lanzamiento del modelo, que está programado para finalizar antes de finales de 2021, intentará ampliar el límite hacia las lenguas con más de un millón de hablantes.

Otros sesgos marginales del modelo pueden resultar de la adopción de estructuras implícitas en las fuentes principales. Por ejemplo, la división en países se derivó de la clasificación de la UIT y no distingue ciertos territorios y, por lo tanto, les asigna el mismo porcentaje que el país de origen (si la tasa de conectividad de un territorio adjunto es de hecho muy inferior a la del país de origen, apego que podría ser el caso de Mayotte, las lenguas específicas de este territorio, por ejemplo, el kibushi, se beneficiarán de un sesgo favorable).

5.2 Sesgo de selección de fuente

Es evidente que existe un "sesgo de selección", que no es específico de la metodología, sino que pertenece a la aplicación del método, donde la decisión sobre la selección de fuentes favorece implícitamente criterios específicos del origen cultural del autor quien inconscientemente ignora países demasiado alejados de su experiencia. Esto puede aplicarse a cada uno de los indicadores y tener un impacto específico en el indicador de tráfico donde la selección de sitios web tiene influencia alta, incluso si el número de sitios web es de cientos. Se implementó el uso de la media truncada al 20% para reducir dichos sesgos, luego de verificar que el 20% era un intervalo grande capaz de eliminar la gran mayoría de resultados centrados en sitios web con fuerte localización lingüística.

5.3 Sesgo de fuente

Los sesgos resultantes de las fuentes se analizan en la tabla a continuación, calificando cada indicador con un valor de 0 (totalmente sesgado) a 20 (sin ningún sesgo).

Tabla 11: Evaluación del sesgo por indicador

INDICADOR	VALOR	REALIMENTACIÓN
INTERNAUTAS	19→16	Este indicador se deriva de un único micro-indicador. La fuente principal es la UIT. En 2017, fue la fuente mejor calificada con un 19/20, pero en esta versión la puntuación cae a 16 porque la UIT ha dejado de proporcionar su propia estimación cuando el país no produce datos oficiales. Las cifras de la UIT se han complementado con cifras del Banco Mundial y se ha establecido una proyección lineal de los datos de años anteriores para los demás casos. Este indicador es fundamental en el método porque se utiliza para ponderar los resultados en varias situaciones, sin embargo, el análisis factorial mostró que el impacto de una pequeña variación es moderado. Por ejemplo, si la tasa de personas conectadas para Brasil se estableciera en 80% en lugar del valor real del 74%,

INDEX	15→18	Este indicador se deriva de una combinación de 25 micro-indicadores que evalúan diferentes parámetros nacionales que caracterizan la sociedad de la información. Las fuentes son organizaciones internacionales, ONG o universidades. Los sesgos, si existen, son marginales. El sesgo de selección aquí es extremadamente bajo porque estamos cerca de la exhaustividad para todos los micro-indicadores.
CONTENIDOS	5→8	Solo hay 13 micro-indicadores para construir este indicador y 11 de ellos provienen de Wikimedia. La distribución de contenidos en la web por lengua es un continente oculto de la Internet y las fuentes existentes son extremadamente raras y, con demasiada frecuencia, sesgadas. Desafortunadamente, el modelo no escapa a esta situación en su estado actual. Como se basa en gran medida en las excelentes estadísticas de Wikimedia, el indicador tiene el sesgo de Wikimedia, donde la presencia de lenguas asiáticas es mucho menor que su proporción en la vida real. Obviamente, el sesgo de selección en este caso, que depende en gran medida de las estadísticas de Wikimedia, es extremadamente importante. Se ha puesto en marcha un sistema de ponderación para reducir esta dependencia tanto como sea posible (que en cualquier caso no ciertamente no es suficiente, razón por la cual la calificación pasó de un muy bajo 5 a un insuficiente 8). El sesgo propio del indicador de contenido es bastante sensible (es decir que las variaciones producen fuertes impactos en los resultados) como lo demuestra el experimento realizado jugando con el sistema de ponderación. Algunas ideas para intentar remediar este problema se implementarán en la próxima edición. Mientras, los sesgos se superan "a mano" utilizando ciertas técnicas (consulte Corrección de sesgos).
TRÁFICO	13→11	Este indicador se deriva de la medición del tráfico por país utilizando Alexa.com en una selección de 338 sitios en la Web. En 2017, el análisis de sesgo mostró que esta fuente estaba fuertemente sesgada contra los países asiáticos y Brasil. En 2021, parece que el sesgo contra los países asiáticos se ha corregido (¡quizás demasiado en el caso de India!) pero se detectan nuevos sesgos que ahora perjudican a los países europeos. El sesgo de selección es evidente en este caso y la próxima versión aumentará seriamente el número de sitios medidos. Se debe explorar la posibilidad de fusionar en proporciones iguales los resultados de Semrush y Alexa para contener los sesgos existentes.
INTERFACES	19	Estos son datos objetivos (si una lengua está presente en la interfaz de una aplicación o como destino para un servicio de traducción en línea). Puede existir sesgo de selección y puede ser necesario ampliar la lista, pero su impacto es marginal. Intuitivamente, podemos ver un aumento, en comparación con 2017, en el número de lenguas admitidos en interfaces o traducción; sin embargo, sigue siendo un

		"indicador radical" que omite la gran mayoría de las lenguas del mundo y se centra en un subconjunto muy limitado.
USOS	12	Este indicador se basa principalmente en los datos de suscripción a redes sociales por país. Si bien los datos recopilados pueden considerarse confiables, el método implica un sesgo contra países no occidentales con aplicaciones alternativas a Facebook, Twitter, LinkedIn, etc. La próxima campaña de medición intentará identificar poblaciones alternativas de suscriptores de aplicaciones para equilibrar los resultados y tratar de reducir el sesgo. Durante este tiempo, la corrección de sesgo debe realizarse "a mano". El sesgo de selección no existe realmente porque la selección está dictada por la estrechez de las opciones existentes. La próxima versión se beneficiará de un pequeño presupuesto para la base de datos comerciales que debería permitir ampliar el número de micro-indicadores.

Si se aplica la ponderación presentada en el cuadro anterior a los resultados en la construcción del promedio ponderado de los macro-indicadores de *potencia* (en lugar del promedio simple), de manera que se tenga en cuenta la confianza relativa en los distintos indicadores del modelo, se anotan los siguientes ajustes en los resultados, a comparar con los resultados anteriores, para tenerlos en cuenta a la hora de corregir los sesgos.

Tabla 12 : Indicadores macro para los 15 lenguas principales después de ponderar los indicadores

	POTENCIA.	Capac.	Grad.	POTENCIA.	Capac.	Grad.	Efecto
Inglés	24,23%	1,86	1,58	26,48%	2,04	1,73	---
Chino	15,77%	1.07	0,89	13,92%	0,95	0,79	+++
Español	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	+++
Francés	3,62%	1,40	1,21	3,71%	1,44	1,24	-
Portugués	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	++
Ruso	3,24%	1,30	0,92	3,11%	1,25	0,88	++
Alemán	2,72%	2,08	1,30	2,86%	2.19	1,37	-
Japonés	2,51%	2,06	1,22	2,52%	2,07	1,22	
Malayo	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	
Bengalí	0,91%	0,35	0,79	0,88%	0,34	0,78	+

5.3.1 Sesgos de Wikimedia

Las estadísticas de Wikipedia son impecables, sin embargo, debe entenderse que, aunque es una de las aplicaciones de la Internet más globales, muestra números para algunos idiomas asiáticos que están muy por debajo de su presencia relativa en la red. La siguiente tabla compara las proporciones entre el número de artículos de Wikipedia y el número de Internautas. Aparecen grandes diferencias con valores anormalmente bajos para las lenguas asiáticas (con algunas excepciones notables).

Tabla 13: Ordenado por número de artículos de Wikipedia

Lengua	Artículos	% ARTE TOTAL.	Ponderación %	Art./L1+L2
Inglés	6332139	12,92%	0,28%	7
Cebuano	5853095	11,94%	22,16%	851
Sueco	3050759	6,22%	14,11%	250
Alemán	2593827	5,29%	1,16%	22
Árabe	2433772	4,97%	0,40%	11
Francés	2342875	4,78%	0,53%	14
Holandés	2060512	4,20%	5,10%	92
Chino	1752600	3,58%	0,07%	2
Ruso	1736736	3,54%	0,41%	9
Italiano	1703284	3,47%	1,51%	33
Español	1698331	3,46%	0,19%	4
Serbocroata	1514114	3,09%	4,57%	97
Polaco	1480982	3,02%	2,20%	45
Japonés	1277204	2,61%	0,61%	11
Vietnamita	1266628	2,58%	1,00%	24
Ucranio	1100281	2,24%	2,00%	52
Portugués	1067241	2,18%	0,25%	6
Malayo	936876	1,91%	0,23%	8
Persa	816984	1,67%	0,59%	15
Coreano	543656	1,11%	0,40%	10
Finlandés	512026	1,04%	5,36%	99
Húngaro	489514	1,00%	2,36%	49
Checo	484445	0,99%	2,18%	44
Rumano	421153	0,86%	1,06%	23
Armenio	420677	0,86%	6,60%	156
Azerbaiyano	420677	0,86%	1,06%	24
Turco	410954	0,84%	0,28%	6
Tártaro	299494	0,61%	3,42%	73
Hebreo	298053	0,61%	1,92%	37
Bielorruso	281379	0,57%	4,34%	87
Búlgaro	273163	0,56%	2,00%	47
Danés	267641	0,55%	2,88%	49
Eslovaco	237210	0,48%	1,98%	40
Kazajo	228493	0,47%	1,05%	23
Griego	195481	0,40%	0,89%	19
Urdu	164062	0,33%	0,04%	3
Hindi	148545	0,30%	0,01%	1
Uzbeko	140894	0,29%	0,25%	9
Tamil	138490	0,28%	0,10%	4

Tailandés	137351	0,28%	0,14%	3
Bengalí	109438	0,22%	0,02%	2

Nótese la presencia del cebuano en segunda posición y la presencia relativa del chino y las lenguas de la India. Es útil comprobar un porcentaje ponderado según el número de hablantes L1 + L2: el inglés no parece desproporcionado y algunas lenguas parecen tener una fuerte presencia en relación con su población L1 + L2, en orden de importancia: cebuano, sueco, armenio, finlandés, holandés, serbocroata, bielorruso y tártaro.

Wikimedia es a la vez el espacio virtual con mayor diversidad lingüística y el único que proporciona sistemáticamente estadísticas lingüísticas fiables y claras sobre todas sus actividades. Sumando la importancia central de su función en la Web, es sin duda un indicador imprescindible cuando se trata de *contenidos*. Desafortunadamente, un análisis serio muestra que este espacio tan especial de ninguna manera podría reflejar una verdadera indicación de la distribución de contenidos por lengua en la Web. La importancia de las lenguas en Wikimedia no siempre está ligada a su importancia real en el ciberespacio y algunas lenguas han invertido mucho en este espacio, independientemente de su presencia global en la Web.

Como se explicó anteriormente, la cantidad de artículos no es un indicador confiable porque, para algunas lenguas, se han implementado bots que crearon artículos a partir de traducciones, los cuales posteriormente se quedan sin actualización. Para controlar ese fenómeno, hay que prestar atención al número de editores activos, el número de ediciones y la profundidad, indicador creado para reflejar el grado de actualización de los artículos. Se ha desarrollado una fórmula para integrar estos factores y se ha presentado anteriormente. Los resultados ordenados por esta fórmula y presentados como porcentajes son los siguientes:

Tabla 14: Artículos de Wikipedia ordenados por fórmula

Inglés	53,96%
Hebreo	10,32%
Italiano	6,99%
Serbocroata	6,02%
Alemán	5,70%
Francés	2,96%
Sueco	2,55%
Finlandés	2,39%
Holandés	1,54%
Japonés	0,97%
Polaco	0,87%
Armenio	0,84%
Húngaro	0,77%
Checo	0,63%
Ucranio	0,54%
Danés	0,50%
Ruso	0,48%
Persa	0,40%

Bielorruso	0,29%
Español	0,29%
Portugués	0,18%
Árabe	0,16%
Rumano	0,13%
Búlgaro	0,11%
Coreano	0,10%
Turco	0,10%
Griego	0,07%
Eslovaco	0,04%
Cebuano	0,03%
Azerbaiyano	0,02%
Malayo	0,02%
Tailandés	0,01%
Chino	0,01%
Malayalam	0,00%
Kazajo	0,00%
Africaans	0,00%
Tártaro	0,00%
Bengalí	0,00%
Mongol	0,00%
Tagalo	0,00%

Esta es claramente una representación más justa de la realidad con Wikipedia, prestando atención equilibrada al número de editores, ediciones y profundidades, y ponderada por el número de hablantes L1 + L2. Tenga en cuenta que el cebuano está penalizado con esta fórmula por su política de uso de bots, pero que otra lengua de Filipinas logra colarse en la tabla: el Tagalo. El predominio del inglés en Wikimedia también aparece más claramente con este enfoque.

Wikimedia no se limita a Wikipedia y también existen estadísticas para cada uno de los otros elementos: WikiBooks, WikiQuote, WikiSource, Wikiversity, Wiktionary, WikiNews y WikiVoyages, para los cuales está disponible el número de artículos por lengua. Para estos elementos de Wikimedia, las fuentes se presentan en términos absolutos, sin ponderar según el número de hablantes, mostrando solo los primeros.

Tabla 15: Número de wikilibros

Inglés	3851195	35,72%
Alemán	961696	8,92%
Francés	657991	6,10%
Portugués	473196	4,39%
Italiano	411671	3,82%
Polaco	403336	3,74%
Húngaro	401256	3,72%
Español	396546	3,68%

Holandés	349987	3,25%
Vietnamita	256386	2,38%
Ruso	205469	1,91%
Japonés	178783	1,66%
Árabe	174452	1,62%
Hebreo	164355	1,52%
Chino	141302	1,31%
Finlandés	131314	1,22%
Persa	112964	1,05%
Malayo	89019	0,83%
Hindi	73969	0,69%

Tabla 16: Número de citas (WikiQuote)

Inglés	33897	14,28%
Italiano	30799	12,98%
Polaco	28960	12,20%
Ruso	13148	5,54%
Checo	9263	3,90%
Persa	8495	3,58%
Alemán	7879	3,32%
Portugués	7443	3,14%
Español	7116	3,00%
Serbocroata	7022	2,96%
Francés	5923	2,50%
Ucranio	5798	2,44%
Eslovaco	4547	1,92%
Turco	4503	1,90%
Búlgaro	4389	1,85%
Hebreo	4202	1,77%

Tabla 17: Número de Wikisources

Francés	2609546	25,3%
Inglés	2204231	21,3%
Chino	778716	7,5%
Bengalí	722295	7,0%
Polaco	669381	6,5%
Ruso	642705	6,2%
Alemán	431714	4,2%
Italiano	415032	4,0%
Tamil	411502	4,0%
Hebreo	214947	2,1%
Sueco	84882	0,8%
Árabe	80708	0,8%
Plurilingüe	78809	0,8%
Armenio	75487	0,7%

Portugués	73139	0,7%
-----------	-------	------

Tabla 18: Número de Wikiversidad

Alemán	49011	36,9%
Inglés	38612	29,0%
Francés	17553	13,2%
Ruso	5883	4,4%
Checo	5195	3,9%
Portugués	4692	3,5%
Italiano	4472	3,4%
Español	2662	2,0%
Finlandés	1914	1,4%
Esloveno	1252	0,9%
Sueco	858	0,6%
Griego	644	0,5%
Japonés	207	0,2%

Tabla 19: Número de entradas de Wikcionario

Inglés	5923218	19,2%
Malgache	5466228	17,7%
Francés	3392407	11,0%
Chino	1239843	4,0%
Serbocroata	1177979	3,8%
Ruso	1002462	3,2%
Español	885649	2,9%
Alemán	737337	2,4%
Holandés	686499	2,2%
Sueco	674872	2,2%
Polaco	649612	2,1%
Kurdo	635201	2,1%
Lituano	616313	2,0%
Griego	462897	1,5%
Italiano	434058	1,4%
Coreano	398737	1,3%
Finlandés	374056	1,2%

Es importante tratar de comprender qué sucedió con el malgache y preguntarse si su clasificación anormal en la tercera posición en el macro-indicador de *gradiente* invalida el método. Esta lengua ocupa el segundo lugar en Wikcionario y muestra un porcentaje del total de entradas del 17%, extremadamente desproporcionado en comparación con su población (18 millones de hablantes) y mucho más aun en comparación con su bajísimo número de hablantes conectados (1,8 millones). A pesar de que el peso de este micro-indicador se ha fijado en 0.1

(igual que todos los Wikimedia excepto la fórmula de Wikipedia y Wikilibros) la desproporción es tan grande que afecta a un promedio ponderado con solo 9 ítems y, en cascada, los macro-indicadores *potencia* y *gradiente*.

Tabla 20: Número de Wikinoticias

Inglés	21687	14,9%
Francés	20761	14,3%
Ruso	17649	12,1%
Polaco	14357	9,9%
Español	11312	7,8%
Chino	8559	5,9%
Árabe	7578	5,2%
Serbocroata	5650	3,9%
Checo	5608	3,9%
Catalán	4056	2,8%
Tamil	3363	2,3%
Sueco	3317	2,3%
Griego	3084	2,1%
Ucranio	1738	1,2%
Rumano	1697	1,2%
Persa	1645	1,1%
Búlgaro	1562	1,1%
Portugués	1474	1,0%
Alemán	1386	1,0%

Tabla 21: Número de artículos en Wikivoyages

Inglés	28852	28,1%
Alemán	16545	16,1%
Persa	8674	8,5%
Italiano	7619	7,4%
Francés	7407	7,2%
Polaco	6946	6,8%
Ruso	5438	5,3%
Holandés	3671	3,6%
Portugués	3624	3,5%
Chino	2972	2,9%
Español	2524	2,5%
Hebreo	2072	2,0%
Vietnamita	1624	1,6%
Sueco	1522	1,5%
Griego	1408	1,4%
Rumano	917	0,9%
Ucranio	779	0,8%

La diversidad de los resultados según cada tema hace imposible sacar una conclusión sistemática del análisis de estas cifras, sin embargo, se pueden hacer algunas afirmaciones generales:

- El inglés está generalmente, pero no siempre, en primer lugar, aunque la proporción de inglés es menos predominante que en Wikipedia, y se mantiene entre el 14% y el 36%, con una media del 23,5% (frente al 29,4% de los indicadores de Wikipedia)³³.
- Francés y alemán puntúan alto en la mayoría de los indicadores de Wikimedia.
- Chino, hindi, bengalí y persa se abren paso en algunos de los indicadores.
- Ciertas lenguas inesperadas aparecen en la parte superior de la lista para ciertos indicadores: malgache y tamil (además del cebuano).

En conclusión, Wikimedia sigue siendo, de lejos, el lugar con mayor diversidad lingüística en la Internet, con algunas lenguas minoritarias logrando obtener una puntuación alta, pero difícilmente refleja la diversidad real de contenido en la Web. El inglés es predominante, pero no tanto como solía ser. En todos los casos, el método debe, como prioridad, mejorar la calidad del indicador de *contenidos*, el cual es preocupante en su estado actual. Un enfoque a explorar es el de identificar aplicaciones similares a Wikimedia que han logrado ocupar un nicho en los espacios lingüísticos y de alguna manera introducirlas en la estadística.

5.3.2 Sesgos de Alexa

La siguiente tabla muestra las diferentes pruebas y comparaciones realizadas entre Alexa y Semrush y, para Alexa, entre los dos años de uso (2017 y 2021). Para Alexa 2017, las cifras de tráfico anteriores se insertaron en el modelo 2021 para hacer una comparación justa, no influenciada por los cambios del modelo. La comparación no se realiza sobre la base de los insumos (por país) sino sobre la base de los resultados del modelo (por lengua); es decir, la comparación se realiza con el producto del modelo insertando como entrada cada una de las respectivas cifras de tráfico. Las comparaciones resaltan (en rojo en la tabla) muchas anomalías.

Tabla 22: Comparaciones de diferentes medidas de tráfico

	SEMRUSH 2021	ALEXA 2021	2021 (SA) / S	ALEXA 2017	A21-A17 / A21
Inglés	52,50%	35,83%	32%	45,40%	-27%
Chino	1,88%	7,67%	-308%	4,94%	36%
Español	14,45%	10,14%	30%	7,53%	26%
Francés	4,48%	2,56%	43%	6,35%	-148%
Ruso	1,88%	1,83%	3%	1,68%	8%
Alemán	2,61%	1,33%	49%	2,94%	-122%
Portugués	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%

³³ Estos porcentajes se refieren al número de artículos en inglés dividido por el número total.

Japonés	0,65%	1,94%	-198%	1,90%	2%
Malayo	0,68%	0,98%	-44%	1,23%	-27%
Italiano	0,89%	0,53%	41%	0,91%	-72%
Turco	0,60%	1,03%	-74%		
Polaco	0,47%	0,31%	34%	0,63%	-100%
Coreano	0,50%	0,90%	-78%	0,72%	20%
RESTO	13,95%	25,34%	-82%	18,99%	25%
TOTAL	100,00%	100,00%	0%	100,00%	0%

- 1) Está claro que Semrush y Alexa no reflejan la misma distribución de tráfico por país para el mismo conjunto de sitios web, las discrepancias son significativas en demasiados casos. Alexa corrigió su sesgo negativo con los países asiáticos, sin embargo, esta vez es Semrush quien parece ignorar a los países asiáticos y árabes.
- 2) Al comparar los resultados de Alexa de 2017 a 2021, esperaríamos, por lo tanto, cambios evolutivos relativamente menores. Este no es el caso de las siguientes lenguas: francés, alemán, italiano y polaco cuyas cifras han bajado de manera sospechosa, una confirmación del sentimiento vivido durante las mediciones de que los países europeos fueron subestimados en las cifras de Alexa2021.

Finalmente, estas comparaciones tienden a confirmar situaciones que se tendrán en cuenta a la hora de corregir los sesgos:

- Inglés, español, hindi parecen sobrevalorados
- El francés, el alemán, el italiano y el polaco parecen muy subestimados.
- El portugués y el malayo parecen subestimados

Para la próxima edición, se debe prestar mucha atención a este indicador para tratar de superar la situación, tal vez una fusión de datos de los servicios existentes podría ser una alternativa para compensar los sesgos.

5.4 Corrección de sesgos

En este punto, no se trata de aplicar la corrección de sesgo a todas las lenguas del estudio, sino de centrarse solo en las 15 lenguas principales en términos de *potencia*. En el futuro, sería interesante integrar la corrección de sesgos en el modelo (se dio un primer paso con la ponderación de los indicadores según la confianza).

Existe un método que se utilizó en 2017 para producir una estimación del porcentaje de *contenidos* que se basa en la consistencia del factor de productividad (relación de *contenidos* a población conectada) para cada lengua considerada y, muy importante, para el resto de las lenguas. Este método aplicado en 2021 conduce a la siguiente estimación aproximada:

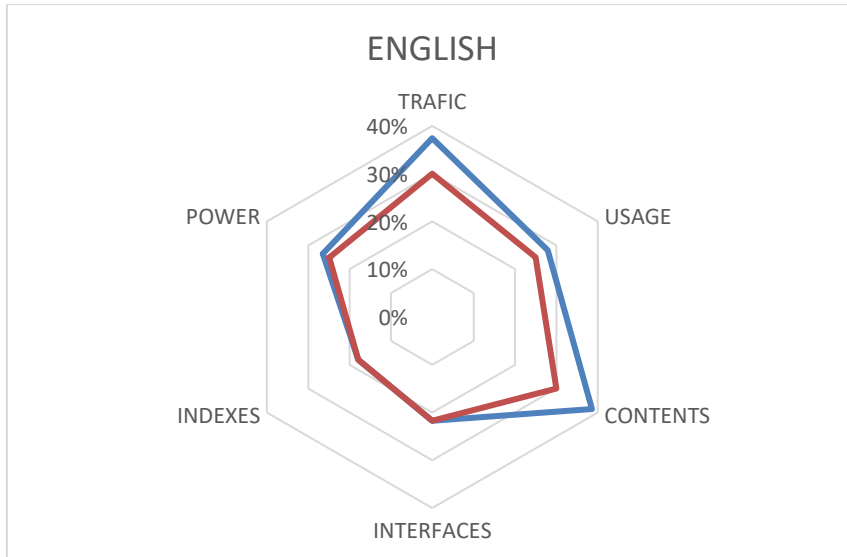
Tabla 23: Primer método de corrección de sesgos

LENGUA	CONTENIDOS	PRODUCTIVIDAD
Inglés	25,0%	1,92
Chino	15,0%	1,02
Español	7,0%	1,34
Francés	4,0%	1,55
Hindi	4,0%	0,69
Portugués	3,5%	1,41
Ruso	3,5%	1,41
Árabe	2,5%	0,71
Alemán	2,5%	1,92
Japonés	2,5%	2,05
Malayo	1,8%	0,76
Italiano	1,4%	2,14
Turco	1,2%	1,41
Coreano	1,2%	1,53
Bengalí	1,2%	0,46
Vietnamita	0,70%	0,94
RESTO	23,00%	0,58

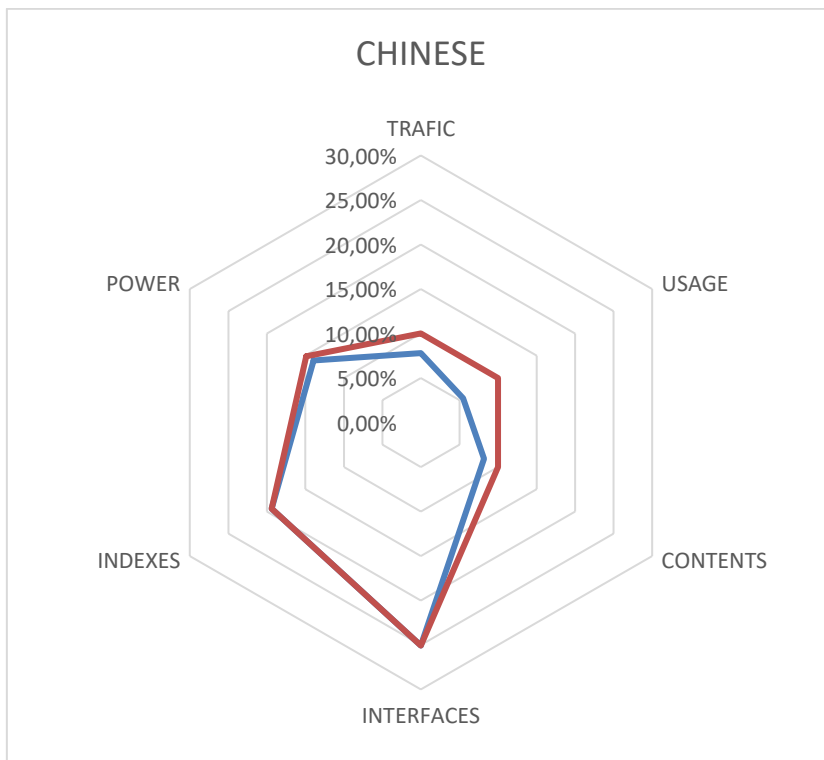
En esta ocasión, se ha agregado un nuevo enfoque para la corrección de sesgos, trabajando específica y directamente en los respectivos sesgos de cada indicador, como se discutió en capítulos anteriores. El patrón de puntuación de lengua se revisa, indicador por indicador, a la luz de lo que se conoce sobre el sesgo, y se registra una posible nueva cifra. A partir de ahí se calcula una nueva cifra de *potencia* con valores redondeados. En azul el resultado del modelo, en rojo los valores corregidos.

Tabla 24: Corrección de sesgo 2do método

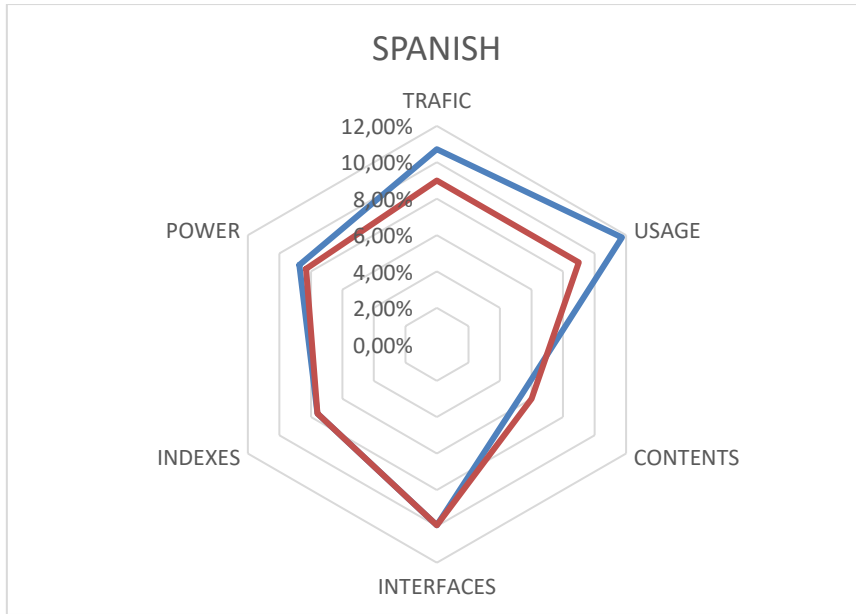
inglés	TRÁFICO	USOS	CONTENIDOS	INTERFACES	INDEXS	POTENCIA
MODELO	37,44%	27,92%	38,61%	21,73%	17,87%	26,48%
CORRECCIÓN DE SESGO	30%	25%	30%	22%	18%	25%



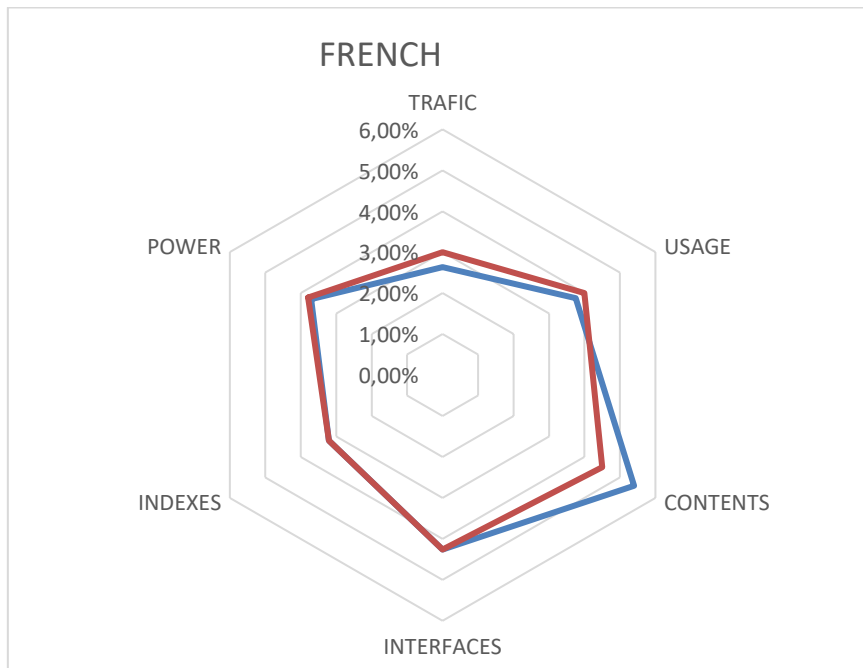
chino	TRÁFICO	USOS	CONTENIDOS	INTERFACES	INDEX	POTENCIA
MODELO	7,79%	5,47%	8,18%	25,07%	19,38%	13,92%
CORRECCIÓN DE SESGO	10%	10%	10%	25%	19%	15%



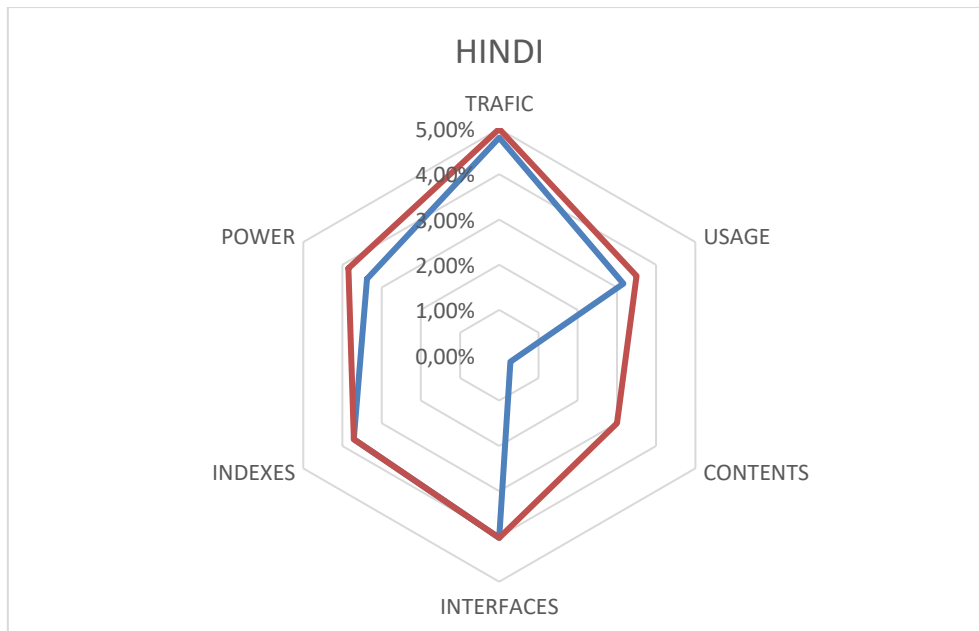
Español	TRÁFICO	USOS	CONTENIDOS	INTERFACES	INDEX	POTENCIA
MODELO	10,72%	11,74%	5,42%	9,94%	7,59%	8,73%
CORRECCIÓN DE SESGO	9%	9%	6%	diez%	8%	8%



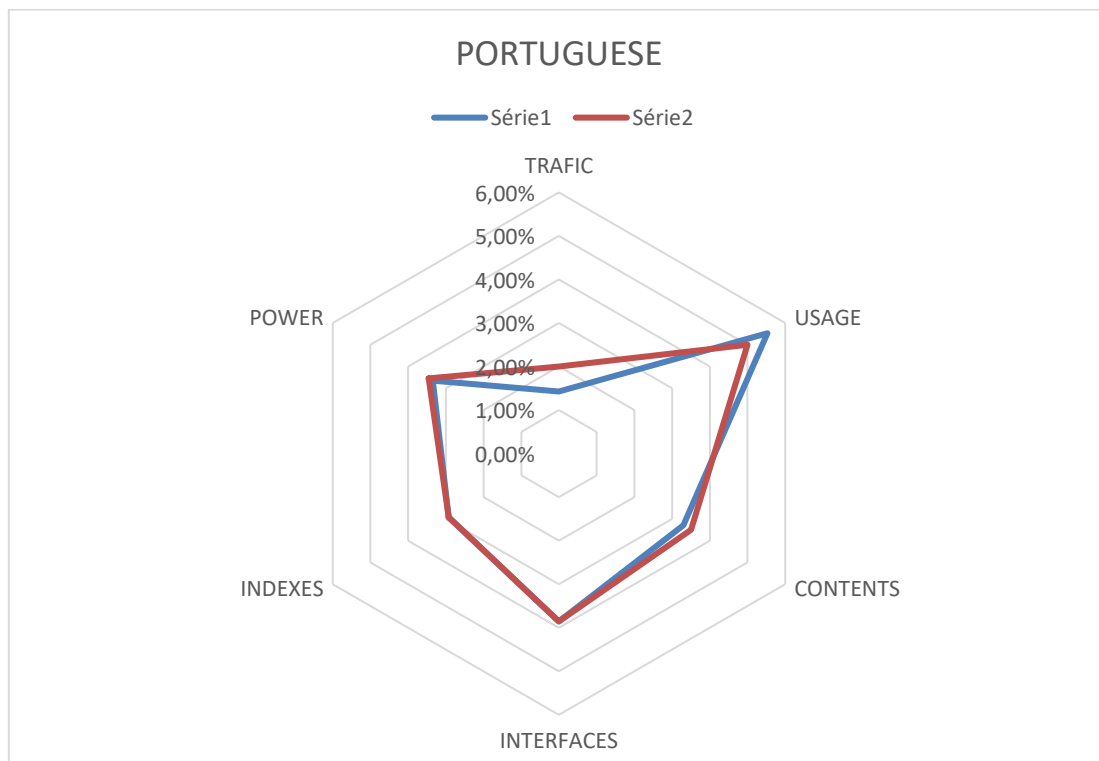
francés	TRÁFICO	USOS	CONTENIDOS	INTERFACES	INDEX	POTENCIA
MODELO	2,64%	3,75%	5,40%	4,26%	3,21%	3,71%
CORRECCIÓN DE SESGO	3,0%	4,0%	4,5%	4,3%	3,2%	3,8%



hindi	TRÁFICO	USOS	CONTENIDOS	INTERFACES	INDEX	POTENCIA
MODELO	4,81%	3,16%	0,28%	4,03%	3,71%	3,38%
CORRECCIÓN DE SESGO	5,0%	3,5%	3,0%	4,0%	3,7%	3,8%



portugués	TRÁFICO	USOS	CONTENIDOS	INTERFACES	INDEXS	POTENCIA
MODELO	1,42%	5,53%	3,30%	3,85%	2,92%	3,35%
CORRECCIÓN DE SESGO	2,0%	5,5%	3%	3,9%	2,9%	3,5%



El resultado de este ejercicio de corrección de sesgo se presenta a continuación y se compara con los resultados del primer método de corrección:

Tabla 25: Resultados de la corrección de sesgo

	SEGUNDO MÉTODO		PRIMER MÉTODO
	POTENCIA	CONTENIDO	POTENCIA
Inglés	25%	30,0%	25%
Chino	15%	10%	15%
Español	8%	6%	7%
Francés	3,8%	4,5%	4%
Hindi	3,8%	3,0%	4%
Portugués	3,5%	2,8%	3,5%

Es interesante constatar que los resultados de los dos métodos diferentes están bastante cerca.

6. CONCLUSIONES Y PERSPECTIVAS

Esta segunda versión del método para producir indicadores de presencia de lenguas en la Internet muestra mejoras interesantes, en particular con datos demo-lingüísticos más fiables y en la gestión de segundas lenguas y multilingüismo. También está avanzando con un enfoque coherente para establecer el porcentaje global del número total de hablantes L1 + L2. Ahora tiene un indicador de índice más completo. El método mejoró el análisis de los sesgos producidos al profundizar el de las estadísticas de Wikimedia y presentando dos formas complementarias para compensar parcialmente estos sesgos.

Sin embargo, el método se enfrenta a importantes desafíos nuevos:

- ✓ con el comportamiento divergente de las herramientas de medición del tráfico;
- ✓ con un indicador de *contenidos* demasiado dependiente de las cifras de Wikimedia que no son geográficamente homogéneas, no reflejan realmente la realidad del contenido y cuya alta sensibilidad ejerce una influencia desproporcionada, en particular sobre el macro-indicador de *gradiente*;
- ✓ con un indicador de *usos* demasiado marcado por las aplicaciones de redes sociales occidentales;
- ✓ y con el hecho de que la UIT ya no proporciona estimaciones para el porcentaje de personas conectadas a la Internet por país (y un problema particular sobre el porcentaje exacto para la India).

Se planea una nueva versión antes de finales de 2021 que intentará abordar estos desafíos y tratar de expandir la cantidad de lenguas procesadas, empujando el límite a las lenguas con más de un millón de hablantes de L1.

El objetivo de la futura versión será también ampliar el número de sitios web medidos en términos de tráfico para potenciar la posibilidad de resultados diferenciados temáticamente más precisos y fiables para determinadas lenguas.

En términos de resultados, continúa la tendencia a la reducción relativa del dominio del inglés con una presencia estimada en torno al 25% (frente al 30% en 2017), el crecimiento del chino y la aparición del hindi como probable cuarta lengua de la Internet, junto al francés hoy, y probablemente por encima en los años venideros.

REFERENCIAS

- [1] D. Pimienta, “Un enfoque alternativo para producir indicadores lingüísticos en la Internet”, 2017 <http://funredes.org/lc2017/Alternativa%20Lengua%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, D. Prado, "Medición de la presencia de la lengua española en la Internet: métodos y resultados", en Revista Española de Documentación Científica 39(3), julio-septiembre 2016, e141 ISSN-L:0210-0614. doi: <http://dx.doi.org/10.3989/redc.2016.3.1328>

ANEXO 1. LISTA DE MICRO INDICADORES Y FUENTES

MICRO-INDICADOR	TIPO	TEMA	URL FUENTE
Amazon EE. UU.: cantidad de libras 2017	CONTENIDOS	Book	Reanudación de 2017
Valor de profundidad de Wikipedia	CONTENIDOS	Ency	https://meta.wikimedia.org/wiki/List_of_Wikipedias
Número de usuarios activos de Wikipedia	CONTENIDOS	Ency	https://meta.wikimedia.org/wiki/List_of_Wikipedias
Número de ediciones de Wikimedia	CONTENIDOS	Ency	https://meta.wikimedia.org/wiki/List_of_Wikipedias
Número de libros wiki por lengua	CONTENIDOS	Book	https://meta.wikimedia.org/wiki/Wikibooks/Table
Número de artículos de Wikipedia por lengua	CONTENIDOS	Ency	https://meta.wikimedia.org/wiki/List_of_Wikipedias
Número de artículos de WikiQuote por lengua	CONTENIDOS	Book	https://stats.wikimedia.org/wikiquote/FR/Sitemap.htm
Número de artículos de WikiSource por lengua	CONTENIDOS	Book	Número de artículos de WikiSource por lengua
Número de artículos de Wikiversity por lengua	CONTENIDOS	S/T	https://stats.wikimedia.org/wikiversity/EN/Sitemap.htm
Número de artículos de Wikcionario por lengua	CONTENIDOS	Dict	https://stats.wikimedia.org/wiktionary/EN/Sitemap.htm
Número de artículos de WikiNews por lengua	CONTENIDOS	News	https://stats.wikimedia.org/wikinews/EN/Sitemap.htm
Número de artículos de WikiVoyages por lengua	CONTENIDOS	Tur	https://stats.wikimedia.org/wikivoyage/EN/Sitemap.htm
Proyección del index T para el comercio electrónico en 2021	CONTENIDOS	e.com	https://translated.com/les-langues-qui-comptent
Index de administración en línea	INDEX	S/T	https://publicadministration.un.org/egovkb/Data-Center*
Index de participación electrónica	INDEX	S/T	https://publicadministration.un.org/egovkb/Data-Center
Index de servicios en línea	INDEX	Infra	https://publicadministration.un.org/egovkb/Data-Center
Index de capital humano	INDEX	ICT	https://publicadministration.un.org/egovkb/Data-Center
Index de infraestructura de telecomunicaciones	INDEX	Gov	https://publicadministration.un.org/egovkb/Data-Center
Index de preparación digital global de Cisco 2019	INDEX	S/T	https://www.cisco.com/c/dam/en_us/about/csr/reports/global-digital-readiness-index.pdf
Index de preparación para la IA del gobierno 2020	INDEX	ICT	https://static1.squarespace.com/static/58b2e92c1e5b6c828058484e/t/5f7747f29ca3c20ecb598f7c/1601653137399/AI+Readiness+Report.pdf
Puntuaciones de libertad en la Internet	INDEX	Book	https://freedomhouse.org/countries/freedom-net/scores
Index de conectividad global	INDEX	Gov	https://www.huawei.com/minisite/gci/en/country-rankings.html
Index de ciberseguridad global 2018	INDEX	Gov	https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-GCI.01-2018-PDF-E.pdf
Index de comercio electrónico B2C de la UNCTAD, 2020	INDEX	Gov	https://unctad.org/system/files/official-document/tn_unctad_ict4d17_en.pdf
El index global de datos abiertos	INDEX	Infra	https://index.okfn.org/place/
Ranking mundial de competitividad digital 2020	INDEX	Secu	https://www.imd.org/globalassets/wcc/docs/release-2020/digital/digital_2020.pdf
Index de preparación para tecnologías fronterizas	INDEX	Econ	https://unctad.org/system/files/official-document/tir2020_en.pdf
Index de innovación global	INDEX	AI	https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2020.pdf
Acceso a conocimientos básicos	INDEX	Econ	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Acceso a información y comunicaciones	INDEX	Gov	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Acceso a la educación superior	INDEX	Gov	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Acceso a electricidad (% de hab.)	INDEX	Infra	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Acceso a una educación de calidad (0 = desigual; 4 = igual)	INDEX	S/T	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Acceso a la gobernanza en línea (0 = bajo; 1 = alto)	INDEX	Econ	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Censura de los medios (0 = frecuente; 4 = raro)	INDEX	Infra	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Libertad de expresión (0 = sin libertad; 1 = plena libertad)	INDEX	Gov	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx

Universidades ponderadas por la calidad (puntos)	INDEX	e.com	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Documentos citables	INDEX	Gov	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Mujeres con educación superior	INDEX	Econ	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Años de educación superior	INDEX	S/T	https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx
Lenguas de traducción de Bing Translator	INTERFAZ	Tra	https://www.bing.com/translator/
Lenguas admitidos por Amazon Kindle Direct Publishing	INTERFAZ	Inter	https://kdp.amazon.com/en_US/help/topic/G200673300
Lenguas admitidos por Cortana	INTERFAZ	Tra	https://en.wikipedia.org/wiki/Cortana
Lenguas de palabras de referencia admitidos	INTERFAZ	Inter	https://www.wordreference.com
Lenguas de traducción de WordLingo	INTERFAZ	Inter	http://www.worldlingo.com/en/languages/
Lenguas admitidos por Facebook	INTERFAZ	Tra	https://www.facebook.com/langue.php
Lenguas de anuncios in-stream de Facebook admitidos	INTERFAZ	Tra	https://www.facebook.com/business/help/267128784014981
Lenguas de traductor gratuitos compatibles	INTERFAZ	Tra	http://www.free-translator.com
Lenguas admitidos por la consola de Google Play	INTERFAZ	Tra	https://support.google.com/googleplay/android-developer/table/4419860?hl=fr
Lenguas admitidos por Google Cloud	INTERFAZ	Inter	https://cloud.google.com/translate/docs/languages?hl=fr
Lenguas admitidos por Google Translate	INTERFAZ	Inter	https://en.wikipedia.org/wiki/Google_Translate
Lenguas admitidos por Google Académico para la búsqueda	INTERFAZ	Inter	https://scholar.google.com/scholar_settings?scifh=1&hl=fr&as_sdt=0,5#1
Lengua admitido por Paralink Translator	INTERFAZ	Inter	http://paralink.com
Lenguas de traducción en línea admitidos	INTERFAZ	Tra	https://www.online-translator.com/traduction
Lenguas compatibles con el traductor Reverso	INTERFAZ	Tra	https://www.reverso.net/text_translation.aspx?lang=FR
Lenguas admitidos por traducción gratuita	INTERFAZ	Tra	https://www.freetranslations.org
Lenguas admitidos por Skype	INTERFAZ	Inter	https://support.skype.com/en/faq/FA34781/what-languages-are-supported-in-skype
Lenguas admitidos por Systran	INTERFAZ	Tra	https://support.systran.net/systranlinks/faq/
163.com	TRÁFICO	GAM	https://www.alexa.com/siteinfo
17ok.com	TRÁFICO	?	https://www.alexa.com/siteinfo
1and1.com	TRÁFICO	Tool	https://www.alexa.com/siteinfo
360.cn	TRÁFICO	Secu	https://www.alexa.com/siteinfo
4shared.com	TRÁFICO	FiSh	https://www.alexa.com/siteinfo
500px.com	TRÁFICO	SN-Im	https://www.alexa.com/siteinfo
6.cn	TRÁFICO	SN-Da	https://www.alexa.com/siteinfo
A2hosting.com	TRÁFICO	Tool	https://www.alexa.com/siteinfo
Abilogic.com	TRÁFICO	DIR	https://www.alexa.com/siteinfo
Acerca de mí	TRÁFICO	Mktg	https://www.alexa.com/siteinfo
Academia.edu	TRÁFICO	S/T	https://www.alexa.com/siteinfo
Adam4Adam.com	TRÁFICO	SN-Da	https://www.alexa.com/siteinfo
Adictinggames.com	TRÁFICO	Gam	https://www.alexa.com/siteinfo
adobe.com	TRÁFICO	ICT	https://www.alexa.com/siteinfo
Adultfriendfinder.com	TRÁFICO	SN-Da	https://www.alexa.com/siteinfo
Aim.com	TRÁFICO	MSG	https://www.alexa.com/siteinfo
Alexa.com	TRÁFICO	Mktg	https://www.alexa.com/siteinfo
Aliexpress.com	TRÁFICO	e.com	https://www.alexa.com/siteinfo
Alipay.com	TRÁFICO	Econ	https://www.alexa.com/siteinfo
Alivedirectory.com	TRÁFICO	DIR	https://www.alexa.com/siteinfo
Amazon.com	TRÁFICO	Book	https://www.alexa.com/siteinfo
Amazonaws.com	TRÁFICO	Host	https://www.alexa.com/siteinfo
Anastasiadate.com	TRÁFICO	SN-Da	https://www.alexa.com/siteinfo
Androide	TRÁFICO	ICT	https://www.alexa.com/siteinfo
Angel.co	TRÁFICO	SN-Fr	https://www.alexa.com/siteinfo
Anobii.com	TRÁFICO	SEng	https://www.alexa.com/siteinfo
Answers.com	TRÁFICO	Q/A	https://www.alexa.com/siteinfo
Aparat.com	TRÁFICO	Vid	https://www.alexa.com/siteinfo
Manzana	TRÁFICO	ICT	https://www.alexa.com/siteinfo
Música de Apple	TRÁFICO	SN-Mu	https://www.alexa.com/siteinfo
Apple.com/Safari	TRÁFICO	ICT	https://www.alexa.com/siteinfo
Archives.org	TRÁFICO	Book	https://www.alexa.com/siteinfo
Archives-ouvertes.fr	TRÁFICO	S/T	https://www.alexa.com/siteinfo
Armorgames.com	TRÁFICO	Gam	https://www.alexa.com/siteinfo
Arvix.com	TRÁFICO	Host	https://www.alexa.com/siteinfo

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

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

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

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

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

Yelp.com	TRÁFICO	SEng	https://www.alexa.com/siteinfo
Youku.com	TRÁFICO	Vid	https://www.alexa.com/siteinfo
YouTube	TRÁFICO	Vid	https://www.alexa.com/siteinfo
Yy.com	TRÁFICO	Vid	https://www.alexa.com/siteinfo
Zhanqi.tv	TRÁFICO	Vid	https://www.alexa.com/siteinfo
Zhihu.com	TRÁFICO	Q/A	https://www.alexa.com/siteinfo
Zillow.com	TRÁFICO	e.com	https://www.alexa.com/siteinfo
Zoho.com	TRÁFICO	Mail	https://www.alexa.com/siteinfo
Zoom.us	TRÁFICO	MSG	https://www.alexa.com/siteinfo
Zoosk.com	TRÁFICO	SN-Da	https://www.alexa.com/siteinfo
% De usuarios de FACEBOOK por país (NapoleonCat 2021)	USOS		https://napoleoncat.com/stats/
INSTAGRAM% de usuarios por país (NapoleonCat 2021)	USOS		https://napoleoncat.com/stats/
MENSAJERO% de usuarios por país (NapoleonCat 2021)	USOS		https://napoleoncat.com/stats/
LINKEDIN% de usuarios por país (NapoleonCat 2021)	USOS		https://napoleoncat.com/stats/
LinkedIn% usuario por país (ApolloTech 2021)	USOS		https://www.apollotechnical.com/linkedin-users-by-country/
Twitter% de usuarios por país (Statista 2021)	USOS		https://www.statista.com/estadísticas/242606/número-de-usuarios-activos-twitter-en-países-seleccionados/
% De usuarios de FACEBOOK de IWS 2021	USOS		https://www.internetworldstats.com/stats1.htm + stats2.htm + ... stats6.htm
% De audiencia de Facebook (Statista 2021)	USOS		https://www.statista.com/estadísticas/268136/top-15-países-basado-en-número-de-facebook-users/
YouTube% de personas conectadas en el país (Statista 2021)	USOS		https://www.statista.com/statistics/1219589/youtube-penetration-worldwide-by-country/
Netflix% de suscriptores por país (CompariTech 2020)	USOS		https://www.comparitech.com/tv-streaming/netflix-subscribers/
% De audiencia de Pinterest (Statista 2021)	USOS		https://www.statista.com/estadísticas/328106/mercados-penetración-de-pinteres/
REDDIT% de usuarios por país (Statista 2021)	USOS		https://backlinko.com/reddit-users
2012/21% de las descargas acumuladas de OpenOffice por país	USOS		http://www.openoffice.org/stats/countries.html
# Servidores de la Internet seguros	USOS		https://data.worldbank.org/indicator/IT.NET.SECR
% Suscripción de banda ancha fija en el país (BM 2021)	USOS		https://data.worldbank.org/indicator/IT.NET.BBND.P2
% Suscripción Tel. Fijo + Móvil en el país (BM 2021)	USOS		https://data.worldbank.org/indicator/IT.MLT.MAIN.P2 + https://data.worldbank.org/indicator/IT.CEL.SETS.P2

TIPOLOGIA	Cant.	TEMA
?	1	
Advo	1	Cabildeo
App	10	Aplicación
Blog	20	Blog
Book	18	Libros
Cloud	3	Nube (informática)
CMS	5	Sistema de gestión de contenidos
DIR	7	Directorio
e.com	9	Comercio electrónico
Econ	5	Economía
EDU	2	Curso
FiSh	11	Compartir archivos
Film	8	Películas
Gam	20	Juegos
Gen	2	Genealogía
Health	2	Salud
Host	7	Alojamiento Web
Hum	1	Humor
ICT	13	TIC

Mail	17	Correo electrónico
Mktg	10	Mercadeo
MOOC	8	MOOC
MSG	23	Mensajería
News	4	Periódicos
Porn	6	Pornografía
Port	8	Portal
Q/A	13	Pregunta/Respuesta
S/T	22	Ciencia y tecnología (investigación)
Secu	2	Seguridad
SEng	26	Motor de búsqueda
SN-Da	20	Redes sociales encuentros
SN-Fr	28	Redes sociales amistad
SN-Im	24	Redes sociales imágenes
SN-Mu	10	Redes sociales música
SN-pr	6	Redes sociales profesionales
Tool	14	Herramientas informáticas
Tur	3	Turismo
VC	1	Comunidades virtuales
Vid	13	Video

ANEXO 2: MACRO-LENGUAS

CÓDIGO ISO	MACRO LENGUA	NÚMERO DE LENGUAS FUSIONADAS
<i>ara</i>	<i>Árabe</i>	29
<i>aym</i>	<i>Aimara</i>	2
<i>aze</i>	<i>Azerbaiyano</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>Estonio</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>Guaraní</i>	5
<i>hai</i>	<i>Haida</i>	2
<i>hbs</i>	<i>Serbocroata</i>	4
<i>hmn</i>	<i>Hmong</i>	25
<i>iku</i>	<i>Inuktítut</i>	2
<i>ipk</i>	<i>Inupiatun</i>	2
<i>jrb</i>	<i>Judeoárabe</i>	5
<i>kau</i>	<i>Kanuri</i>	3
<i>kln</i>	<i>Kalenjin</i>	9
<i>kok</i>	<i>Konkani</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>Kurdo</i>	3
<i>lah</i>	<i>Lahnda</i>	7
<i>lav</i>	<i>Letón</i>	2
<i>luy</i>	<i>Luyia</i>	14
<i>man</i>	<i>Mandingo</i>	6
<i>mlg</i>	<i>Madagascari</i>	11
<i>mon</i>	<i>Mongol</i>	3
<i>msa</i>	<i>Malayo</i>	36
<i>mwr</i>	<i>Marwari</i>	6
<i>nep</i>	<i>Nepalés</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>Pashto</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>Albanés</i>	4
<i>srd</i>	<i>Sardo</i>	4
<i>swa</i>	<i>Swahili</i>	2
<i>syr</i>	<i>Siríaco</i>	2
<i>tmh</i>	<i>Tamasheq</i>	4
<i>uzb</i>	<i>Uzbeko</i>	2
<i>vid</i>	<i>Yídish</i>	2
<i>zap</i>	<i>Zapoteca</i>	57
<i>zha</i>	<i>Zhuang</i>	16
<i>zho</i>	<i>Chino</i>	15
<i>zza</i>	<i>Dimli</i>	2

ANEXO 3: LISTA DE PAÍSES O TERRITORIOS DONDE LA UIT NO OFRECE DATOS

Código ISO	NOMBRE DEL PAÍS	POBLACIÓN
AX	Islas Åland	27.652
AS	Samoa Americana	55.990
IO	Territorio Británico del Océano Índico	4.000
BQ	Caribe neerlandés	18.740
CX	Isla de Navidad	1.170
CC	Islas Cocos (Keeling)	630
CK	Islas Cook	15.000
CW	Curazao	140.000
GF	Guayana Francesa	366.590
GP	Guadalupe	454.800
GU	Guam	139.550
IM	Isla de Man	88.085
MQ	Martinica	377.100
NF	Isla Norfolk	1.500
KP	<i>Corea del Norte</i>	<i>25.579.000</i>
MP	Islas Marianas del Norte	53.280
PW	Palau	17.550
PN	Pitcairn	36
RE	Reunión	751.580
BL	San Bartolomé	7.850
MF	San Martín	28.500
PM	San Pedro y Miquelón	6.340
SX	San Martín	33.470
TC	Islas Turcas y Caicos	30.170
VA	<i>Estado del Vaticano</i>	<i>330</i>
EH	<i>Sahara Occidental</i>	<i>544.150</i>
	TOTAL	28.689.463

Hay dos posibles razones por las que el país o territorio está excluido de los datos de la UIT:

- 1) Es un territorio cuyos datos están incluidos en un país determinado.
- 2) No hay fuente ni estimación del porcentaje de personas conectadas a la Internet (en cursiva en la tabla).

ANEXO 4: RESULTADOS PARA TODAS LAS LENGUAS

Rango			Conect.L.M.	Pop.M.	TRÁF.	Conect.L.	USOD	CONT.	INTERF.	INDEX	POT.	CAPAC.	GRAD ..
	ISO	Total o promedio --->	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	Africaans	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	Assamese	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	Bielorrusia	0,06%	0,04%	0,02%	82,27%	0,03%	0,03%	0,03%	0,06%	0,04%	1,00	0,66
15	ben	Bengalí	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	Checo	0,19%	0,13%	0,07%	81,37%	0,13%	0,50%	0,18%	0,25%	0,22%	1,70	1,14
55	dan	Danés	0,10%	0,05%	0,04%	97,82%	0,08%	0,26%	0,08%	0,16%	0,12%	2,19	1,22
9	deu	Alemán	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	Griego	0,18%	0,13%	0,21%	77,71%	0,17%	0,37%	0,19%	0,24%	0,22%	1,75	1,23
1	eng	Inglés	15,30%	13,01%	37,4%	64,33%	27,9%	38,61%	21,73%	17,87%	26,48%	2,04	1,73
125	ewe	Éwé	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	Finlandés	0,09%	0,06%	0,04%	89,67%	0,06%	0,74%	0,08%	0,14%	0,19%	3,42	2,09
4	fra	Francés	3,00%	2,58%	2,64%	63,67%	3,75%	5,40%	4,26%	3,21%	3,71%	1,44	1,24
70	ful	Fuluide	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	Alemán. Suizo	0,10%	0,06%	0,08%	91,56%	0,09%	0,00%	0,01%	0,17%	0,08%	1,21	0,72
28	guj	Gujarati	0,44%	0,60%	0,53%	40,49%	0,35%	0,05%	0,24%	0,39%	0,34%	0,56	0,76
91	hat	Criollo haitiano	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	Serbo-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	Hebreo	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	Armenio	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	Javanés	0,58%	0,66%	0,20%	47,74%	0,69%	0,00%	0,14%	0,51%	0,35%	0,53	0,61
10	jpn	Japonés	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	Kannada	0,42%	0,57%	0,47%	40,12%	0,31%	0,08%	0,23%	0,36%	0,31%	0,55	0,75
104	kas	Cachemira	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	Kazaj	0,18%	0,13%	0,07%	76,98%	0,10%	0,07%	0,10%	0,17%	0,11%	0,90	0,64
64	khm	Khmer	0,14%	0,17%	0,07%	43,40%	0,16%	0,02%	0,08%	0,09%	0,09%	0,53	0,66
121	kik	Gikuyu	0,03%	0,08%	0,01%	22,57%	0,03%	0,00%	0,01%	0,03%	0,02%	0,22	0,53
111	kin	Kinyarwanda	0,06%	0,13%	0,02%	24,69%	0,02%	0,00%	0,01%	0,04%	0,02%	0,19	0,42
132	kln	Kalenjin	0,02%	0,04%	0,01%	22,62%	0,02%	0,00%	0,00%	0,01%	0,01%	0,21	0,50
137	kmb	Kimbundu	0,00%	0,02%	0,00%	16,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,14	0,48
108	kok	Konkani	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	Kurdo	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	Madura	0,07%	0,08%	0,02%	47,70%	0,08%	0,00%	0,00%	0,06%	0,04%	0,50	0,57
75	mag	Magahi	0,15%	0,20%	0,16%	39,99%	0,11%	0,00%	0,00%	0,13%	0,09%	0,45	0,62
51	mai	Maithili	0,24%	0,33%	0,25%	39,28%	0,18%	0,00%	0,02%	0,20%	0,15%	0,44	0,62
35	mal	Malayalam	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
23	mar	Marathi	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	Malgacheí	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	Malayo	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	Birmanio	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-cal.	0,07%	0,06%	0,03%	74,39%	0,08%	0,00%	0,00%	0,10%	0,05%	0,84	0,62
58	nep	Nepalés	0,16%	0,25%	0,09%	35,70%	0,14%	0,03%	0,14%	0,11%	0,11%	0,45	0,69
22	nld	Holandés	0,40%	0,24%	0,19%	92,02%	0,42%	1,13%	0,47%	0,60%	0,53%	2,26	1,34
90	nod	Tailandés 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 Este	0,33%	0,50%	0,44%	35,80%	0,30%	0,00%	0,03%	0,27%	0,23%	0,45	0,69
17	pol	Polaco	0,58%	0,39%	0,31%	81,17%	0,53%	1,57%	0,69%	0,73%	0,74%	1,88	1,26

6	por	Portugués	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	Rumano	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	Ruso	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	Cingaleses	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	Sindhi	0,11%	0,32%	0,15%	18,73%	0,10%	0,01%	0,03%	0,06%	0,08%	0,24	0,70
98	som	Somalí	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. Del sur	0,13%	0,13%	0,06%	56,47%	0,08%	0,00%	0,01%	0,12%	0,07%	0,51	0,49
105	sou	Tailandés Sur	0,05%	0,04%	0,02%	66,68%	0,06%	0,00%	0,00%	0,05%	0,03%	0,70	0,57
3	spa	Español	7,00%	5,24%	10,7%	73,08%	11,7%	5,42%	9,94%	7,59%	8,73%	1,67	1,25
80	sqi	Albanés	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	Sonde	0,27%	0,31%	0,09%	47,69%	0,33%	0,01%	0,06%	0,24%	0,17%	0,54	0,62
46	swa	Swahili	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	Tamil	0,62%	0,82%	0,71%	41,35%	0,51%	0,19%	0,39%	0,55%	0,50%	0,60	0,80
87	tat	Tatar	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	Tajik	0,05%	0,08%	0,02%	32,22%	0,03%	0,00%	0,01%	0,03%	0,02%	0,29	0,49
33	tgl	Tagalog	0,24%	0,25%	0,33%	53,60%	0,43%	0,06%	0,15%	0,24%	0,24%	0,98	1,00
21	tha	Tailandés	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	Tailandes 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	Turkmen	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	Ucranio	0,37%	0,32%	0,17%	63,96%	0,25%	0,26%	0,33%	0,40%	0,30%	0,92	0,79
131	umb	Umbundu	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	Uzbeko	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 occ.	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	Yoruba	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	Chino	17,65%	14,72%	7,79%	65,59%	5,47%	8,18%	25,07%	19,38%	13,92%	0,95	0,79
48	zul	Zulú	0,29%	0,27%	0,13%	59,57%	0,17%	0,03%	0,09%	0,27%	0,16%	0,60	0,55