



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

<https://OBDILCI.ORG/>

LECCIONES APRENDIDAS DE LA EXPERIENCIA DE MECILDI

INFORME TÉCNICO N.º 1

La desértica web de las lenguas indígenas, vernáculas, locales o minoritarias.

OBDILCI, septiembre de 2026

Versión 2: complementada con .africa, .lat y .in

Tras aplicar el marco computacional MECILDI, sensible al multilingüismo, para medir la distribución lingüística del contenido web en 30 series diferentes (gTLD, ccTLD o Tranco), ha llegado el momento de extraer conclusiones de dicha experiencia. El presente informe técnico es el primero de esta nueva serie de informes técnicos elaborados por el Observatorio de la Diversidad Lingüística y Cultural en la Internet (OBDILCI).

El marco computacional desarrollado por OBDILCI, que mide la distribución lingüística y el multilingüismo en cualquier conjunto de sitios web, se ha aplicado hasta ahora a la serie Tranco de los millones de sitios web más visitados, a 8 gTLD asociados a lenguas de Francia, a 10 ccTLD de países miembros de la Organización de la Francofonía (OIF) y a otros 11 ccTLD adicionales, seleccionados por tener lenguas locales con importantes comunidades de hablantes. Estas mediciones han permitido confirmar algunas cifras previas sobre la distribución lingüística en la web y calcular nuevos indicadores de multilingüismo.

Este primer estudio se centra en los resultados para lenguas **indígenas, vernáculas, locales o minoritarias** y ofrece un panorama particularmente preocupante sobre la escasa presencia de estas lenguas en la web, con datos muy inferiores a las previsiones más pesimistas. En adelante, utilizaremos el término «local» para referirnos a cualquiera de las categorías lingüísticas mencionadas.

La siguiente tabla ofrece la recopilación de datos obtenidos de la ejecución del experimento MECILDI, completada con los indicadores apropiados para establecer comparaciones y clasificaciones ajustadas.

La columna de la tabla contiene las siguientes definiciones:

LENGUAS: el nombre en español de la lengua local

ISO: El código ISO-639-3 asociado a esta lengua.

TLD: El dominio de nivel superior medido por MECILDI.

LUGAR: El lugar (país o región) correspondiente a este TLD.

%SP: Porcentaje de hablantes de L1+L2 de esta lengua en el lugar. (Fuente: Ethnologue)

L1+L2: Número de hablantes de esa lengua en el lugar o, si se indica de otro modo, en el mundo o un país específico. (Fuente: Ethnologue)

%P: % de páginas web en esta lengua entre el total de páginas web en la zona TLD. (MECILDI)

%S: % de sitios web que tienen una configuración lingüística en esa lengua. (MECILDI)

WRate: La relación entre %SP sobre %S, indicador de prevalencia, donde 1 representa una proporcionalidad estricta.

La tabla está ordenada por WRate descendente.

LENGUAS	ISO	TLD	LUGAR	%SP	L1+L2	%P	%S	WRate	N
Vasco	eus	.eus	País Vasco	36%	1M	36,37%	94,02%	2.612	¹
Catalán	cat	.ad	Andorra	37%	30900	20,75%	61,68%	1.672	²
Bretón	bre	.bzh	Bretaña	5%	200K (M)	2,91%	2,19%	0,438	
Papiamentu	pap	.cw	Curazao	92%	136K	11,52%	18,47%	0,201	
Tahitiano	tah	.pf	Polinesia Francesa	19%	56K	1,69%	2,67%	0,144	
Malgache	mlg	.mg	Madagascar	58%	18M	4,61%	7,58%	0,131	
Corso	cos	.corsica	Córcega	35%	120K (M)	3,55%	3,68%	0,105	
Criollo haitiano	hat	.ht	Haití	95%	13,8 M (O)	3,41%	7,89%	0,083	³
Bislama	bis	.vu	Vanuatu	34%	115K (M)	0,99%	1,60%	0,047	
Swahili	swc	.cd	Rep. Dem. del Congo	25%	32M	-	0,53%	0,021	⁴
Yídish	yid	.il	Israel	5%	443K (M)	0,02%	0,03%	0,006	
Rundi	run	.bi	Burundi	98%	13M (M)	-	0,56%	0,006	
Guaraní	grn	.py	Paraguay	88%	6M	0,13%	0,26%	0,003	⁵
Quechua	que	.pe	Perú	21%	7,3M	0,04%	0,06%	0,003	
Criollo de Guadalupe	gcf	.gp	Guadalupe	90%	400K (M)	0,07%	0,18%	0,002	⁶
Tamazight	zgh	.ma	Marruecos	25%	9M	-	0,04%	0,002	
Kabuverdianu	kea	.cv	Cabo Verde	82%	492K	ε	ε	0.000	⁷
Ladino	lad	.il	Israel	0,5%	43K	ε	ε	0.000	⁸
Alsaciano	gsw	.alsace	Alsacia	35%	900K (FR)	0	0	0	
Aymara	aym	.bo	Bolivia	14%	1,7M	0	0	0	

Jula	dyu	.ci	Costa de Marfil	27%	8,5M	0	0	0	
Lenguas kanak		.nc	Nueva Caledonia	36%	71K	0	0	0	9
Kimbundu + Varios		.ao	Angola	90%	33M	0	0	0	10

NOTAS

- 1: Bilingüismo casi perfecto con el español.
- 2: Una distribución bastante equilibrada entre catalán, inglés, español y francés.
- 3: El 80% de la población solo habla criollo.
- 4: Entre muchas otras lenguas indígenas como Lingala (lin) 40M, todas ausentes
- 5: El guaraní es el idioma oficial de Paraguay
- 6: Lo que se detectó es criollo haitiano.
- 7: Se encontró un sitio único entre más de 1100 procesados
- 8: Cuatro sitios web que tienen ladino (lad) donde se encontraron
- 9: Drehu (12K), wallisiano (18K) y otras 32 lenguas kanak.
- 10: Kimbundu 3,7 millones, Umbundu 5,9 millones, Chokwe 2,4 millones, Kikongo 2,4 millones, Nhaneka 1,4 millones, Nganguela 700 mil, Kwanhama 1 millón, Fiote 391 mil, Muhumbi 253 mil, Luvale 180 mil

¿Qué nos dicen esos datos?

Excepto en los casos del euskera y el catalán, que muestran una vitalidad impresionante en el ecosistema web, todas las lenguas locales estudiadas indican una presencia y vitalidad extremadamente bajas en la web. El WRate, tomado como indicador de la prevalencia de la lengua local dentro del TLD correspondiente, muestra valores nulos o casi nulos en un tercio de las lenguas locales, valores críticos, inferiores a 0,01, en más de la mitad de los casos, y valores muy bajos (entre 0,1 y 0,5) para el 20 % de las lenguas locales.

Es evidente que el análisis de unos treinta lugares no es suficiente para extraer conclusiones estadísticas definitivas; sin embargo, sí basta para suscitar una preocupación legítima.

OBDILCI, basándose en sus estudios previos, esperaba encontrar la mayoría de los resultados dentro de la categoría de "valores muy bajos", no en los valores críticos y nulos.

Ejemplos como los siguientes ponen de manifiesto la magnitud del problema al que nos enfrentamos y la urgente necesidad de políticas públicas adecuadas:

- **El guaraní** es idioma oficial de Paraguay, recibe cada vez más atención por parte de las políticas públicas, lo habla el 88% de la población paraguaya, sin embargo, solo el 0,26% de los sitios web tienen una configuración en guaraní, lo que significa pocas unidades.
- En Haití, el 80% de la población solo entiende **el criollo haitiano**, y esa enorme proporción de la población haitiana tiene acceso lingüístico a menos del 8% del ecosistema web nacional.

- Un científico boliviano ha creado y desarrollado, desde finales de los años 80, un algoritmo de traducción notable y bastante original que utiliza **el aimara** como lengua pivote¹; sin embargo, no existe absolutamente ningún sitio web en aimara, a pesar de que lo habla el 14% de la población.
- **El papiamento** es el idioma dominante de la isla de Curazao, notablemente multilingüe; más del 90% de la población lo usa a diario, sin embargo, la web de Curazao apenas contiene poco más del 10% de sus páginas en ese idioma.
- **El kabuverdianu** es el idioma dominante de Cabo Verde, hablado por más del 80% de la población, frente al 60% que habla portugués; sin embargo, el único sitio web identificado pertenece al Instituto del Patrimonio Cultural.
- **El rundi** es el idioma principal de Burundi, entendido por prácticamente todos los habitantes del país, sin embargo, está presente en menos del 1% de los sitios web del dominio .bi.

Nos encontramos cerca de la mitad del *Decenio de las Lenguas Indígenas* promovido por la UNESCO; muchas organizaciones, como la OIF, están organizando talleres de expertos para afrontar los retos que plantea la IA para integrar plenamente las lenguas locales; un número creciente de lenguas (quizás 750) han recibido la atención de UNICODE para obtener la codificación necesaria para su existencia digital, pero ¿qué ocurre con los **contenidos** en esas lenguas?

MÉTODO

Aplicamos el marco MECILDI a la lista completa de sitios web identificados bajo el TLD de los países o regiones mencionados. En algunos casos, utilizamos la zona oficial del TLD, cuando la comparte el centro de información de la red, a través de ICANN/CZDS; en los demás casos, utilizamos un proveedor comercial (zonestats.io). En este último caso, la lista de dominios podría haber omitido algunos sitios web recientes con idiomas locales, pero el resultado estadístico habría sido muy similar.

MECILDI explora sistemáticamente todos los dominios por debajo del TLD; sin embargo, una parte significativa queda excluida de la medición por diversas razones, entre ellas la voluntad del propietario del sitio web de no ser procesado por robots. No obstante, esta situación no alteraría los resultados estadísticos de manera significativa.

Todos los resultados detallados son de consulta pública (véanse las referencias), excepto la última serie de ccTLD, que aún no se han publicado. El ejemplo de Cabo Verde (.cv) se proporciona en el anexo para facilitar la comprensión del proceso y el tipo de indicador obtenido.

¹ <https://www.obdilci.org/blog/ivan-guzman-de-rojas-and-the-aymara-language-3/>

DATOS ADICIONALES: .africa , .lat y .in

Dados los primeros resultados, se han realizado tres mediciones adicionales para confirmar la tendencia observada. Los tres nuevos dominios a medir son:

.africa , una iniciativa gTLD exitosa con más de 60 000 dominios (probablemente el segundo TLD más grande relacionado con África, por debajo de .za (más de 300 000 dominios) pero superior a .ma (30 000), dedicado a sitios web con cobertura panafricana.

.lat , otro gTLD exitoso con más de 300 000 dominios, menos que .br (1,2 millones de dominios) pero más que .cl (menos de 300 000 dominios), dedicado a sitios web con enfoque regional en Latinoamérica y el Caribe. Este gTLD regional geográfico se creó para albergar entidades regionales en la web.

.in, el dominio de nivel superior de código de país (ccTLD) de la India, un importante centro de sitios web en Asia, después de China, en un país con muchos idiomas indígenas ampliamente hablados y, al mismo tiempo, el segundo país del mundo con más hablantes de inglés como lengua materna o segunda lengua, después de Estados Unidos.

Lamentablemente, esas tres mediciones adicionales confirman, en gran medida, las observaciones realizadas hasta ahora sobre la escasez de páginas web en idiomas locales; también confirman observaciones previas sobre el multilingüismo web, según las cuales los países de habla inglesa y los grandes países asiáticos tienen un rendimiento muy bajo en términos de multilingüismo web.

Resultados de .africa

Consulte el anexo para obtener la definición de cada indicador.

Número de sitios web analizados:	59 628
Número de sitios web válidos para el proceso	: 18 181
MultiR	1.50
%Multiplicador 10,70%	
AvgL	5,66
GooT	0,57%
IDN	0,00%

TABLA DE RESULTADOS PARA LOS IDIOMAS MÁS PRESENTES

	%PÁGINAS	%SITIOS
TOTAL	100,00%	149,80%
Inglés	59,74%	89,50%
Francés	8,81%	13,21%
Alemán	3,36%	5,03%
Español	2,04%	3,05%

Portugués	1,82%	2,73%
Vietnamita	1,47%	2,20%
Italiano	1,31%	1,96%
Árabe	1,19%	1,78%
Holandés	1,15%	1,72%
Indonesio	1,04%	1,55%
Chino	0,98%	1,47%
Polaco	0,91%	1,36%
Ruso	0,78%	1,16%
Japonés	0,77%	1,16%
Swahili	0,77%	1,15%
Sueco	0,69%	1,03%

TABLA DE RESULTADOS PARA LENGUAS AFRICANAS

	%PÁGINAS	%SITIOS
Árabe	1,19%	1,78%
Swahili	0,77%	1,15%
Africaans	0,34%	0,51%
Zulú	0,17%	0,25%
Hausa	0,16%	0,24%
Amárico	0,14%	0,21%
Xhosa	0,10%	0,15%
Yoruba	0,10%	0,15%
Igbo	0,09%	0,14%
Kinyarwanda	0,09%	0,14%
Oromo	0,06%	0,08%
Somalí	0,06%	0,08%
Sotho del sur	0,05%	0,07%
Swati	0,05%	0,07%
Tswana	0,05%	0,07%
Lingala	0,04%	0,06%
Tongano	0,04%	0,06%
Vender	0,04%	0,06%
Piel	0,03%	0,04%
Ndebele del sur	0,03%	0,04%
Multitud	0,03%	0,04%
Wolof	0,03%	0,04%
Madagascari	0,02%	0,03%
Shona	0,02%	0,03%
Tigrinya	0,02%	0,03%

El dominio de nivel superior genérico (gTLD) .africa muestra un bajo porcentaje de sitios web multilingües (10,40 % frente a un promedio mundial del 14,4 %), con una cantidad relativamente alta de idiomas por sitio web multilingüe, lo que da como resultado una tasa promedio de multilingüismo. Prácticamente no se utiliza la capacidad de los dominios internacionales (IDN).

La proporción de sitios web con versión en inglés con respecto a los sitios web con versión en francés es mucho mayor que la proporción de hablantes de inglés como L1+L2 en África frente a los hablantes de francés (6,78 frente a 1,62), lo que indica una menor aceptación en los países africanos francófonos.

Solo dos idiomas, el árabe y el suajili, superan el 1% de los sitios web; solo diez idiomas superan el 0,1%; el resto, la mayoría, tienen pocas apariciones en sitios web (oromo, 15 sitios; tigrina, 2). Cabe destacar que el rendimiento del vietnamita y el indonesio se obtiene mediante redireccionamientos a sitios web de juegos globales. También es importante señalar un porcentaje muy alto de dominios no accesibles o que no permiten la detección de idioma (texto insuficiente para el análisis). El alto valor de AvgL se debe a un alto porcentaje de sitios web redirigidos que alcanzan sitios web de comercio electrónico globales. La próxima versión de MECILDI rechazará la duplicación de sitios web redirigidos y probablemente mostrará un valor de AvgL menor, con un impacto marginal en los resultados de los idiomas locales.

Finalmente, cabe señalar que es posible que existieran más lenguas africanas no identificadas por el algoritmo de detección de idiomas, limitado a 220, pero su presencia no superaría las cifras indicadas por OtherL o NON-ID. Por lo tanto, permanecerían en un nivel muy marginal. El análisis de los sitios web clasificados con la etiqueta OtherL muestra una gran proporción de redirecciones a sitios web multilingües globales con un elevado número de idiomas no específicamente vinculados a África.

Resultados de .lat

Los resultados para .lat no se ajustaron sustancialmente a lo esperado para un dominio geográfico regional. De hecho, el dominio se revendió en 2022 a la empresa XYZ, que parece haberlo convertido en un dominio global en lugar del inicialmente dedicado a Latinoamérica y el Caribe. En cualquier caso, la presencia de lenguas indígenas de Latinoamérica y el Caribe es prácticamente nula.

MultiR	1.60
%Multiplicador	15,27%
Promedio L	4,89
GooT	0,16%
IDN	0,06%

El porcentaje de sitios web multilingües es aceptable, superior a la media mundial.

	%PÁGINAS	%SITIOS
	100,00%	159,71%
Inglés	29,40%	45,39%
Indonesio	14,06%	21,69%
Español	13,68%	21,28%
Portugués	6,83%	10,90%
Chino	5,24%	8,40%
Vietnamita	3,44%	5,25%
Ruso	3,03%	5,04%
Bengalí	2,89%	4,50%
Francés	2,83%	4,82%
Alemán	2,59%	4,43%
Turco	2,01%	3,31%
Hindi	1,97%	3,40%
Italiano	1,62%	2,84%
Árabe	1,35%	2,41%
Japonés	1,27%	2,28%
Coreano	1,21%	2,18%
Serbio	1,02%	1,82%
Tailandés	0,71%	1,16%

El inglés está presente en algo menos del 50% de los sitios y el indonesio es un segundo idioma sorprendente, uno de los muchos idiomas asiáticos inesperados (vietnamita, bengalí, hindi). La razón en la mayoría de los casos es que los sitios web de juegos redirigen desde el dominio .lat.

En cuanto a las lenguas indígenas de América Latina y el Caribe, están prácticamente ausentes de los resultados, salvo el criollo haitiano, que aparece casi al final con un 0,03 % de las páginas y un 0,05 % de los sitios web con configuración lingüística en criollo. Fuimos un paso más allá y comprobamos directamente la presencia de los códigos ISO de las lenguas indígenas en el campo Lang= o Hreflang=. Los resultados son totalmente marginales: en total, 7 sitios de entre 100 000 representan un 0,007 %.

IDIOMA	Códigos ISO principales	Otros códigos ISO	Ocurrencias en Lang=	Ocurrencias en HrefLang =
Quechua	que	qul,qva,qve,qvm	3	0
Guaraní	verde / gn	gnw,gui	3	0
Maya de Yucatán	yua / mi		0	0
Aimara	aym /ay	ayc,ayr	0	0
Náhuatl	nhn /nah	nch,ncx	1	0

Resultados para .in

Este es uno de los resultados más esperados del OBDILCI, basado en un informe técnico de KPMG de 2019² que ofrece una perspectiva positiva sobre la presencia de las lenguas locales indias más habladas, en primer lugar, el hindi, la tercera lengua más hablada del mundo.

Lamentablemente, salvo por el hindi, que muestra cierto nivel de presencia, aunque no en la proporción esperada, el lugar de las lenguas indígenas en el ecosistema web de la India sigue siendo muy bajo, al igual que el porcentaje de sitios multilingües.

MultiR	1.28
%Multi	6,52%
PromedioL	5.39
GooT	3,20%
IDN	0,00%

	%PÁGINAS	%SITIOS
	100,00%	128,45%
Inglés	73,00%	91,83%
Hindi	3,81%	4,83%
Vietnamita	2,04%	2,61%
Indonesio	1,53%	1,96%
Español	1,33%	1,77%
Alemán	1,30%	1,74%
Francés	1,19%	1,60%
Portugués	1,04%	1,39%
Italiano	0,91%	1,23%
Tamil	0,79%	1,05%
Japonés	0,70%	0,94%
Chino	0,58%	0,79%
Marathi	0,55%	0,71%
Ruso	0,55%	0,73%
Holandés	0,53%	0,75%
Árabe	0,52%	0,70%
Polaco	0,44%	0,62%
Coreano	0,41%	0,56%
Malayalam	0,40%	0,53%

² <https://assets.kpmg.com/content/dam/kpmgsites/in/pdf/2017/04/Indian-languages-Defining-Indias-Internet.pdf>

El inglés es extremadamente dominante, presente en más del 90% de los sitios web. El hindi ocupa un segundo lugar relevante, aunque con menos del 5%. El vietnamita y el indonesio ocupan los siguientes puestos, muy probablemente debido a las redirecciones hacia sitios web de juegos.

En cuanto a los idiomas de la India, a continuación, se muestra una tabla con la presencia marginal de los más importantes.

	%PÁGINAS	%SITIOS
Hindi	3,81%	4,83%
Tamil	0,79%	1,05%
Marathi	0,55%	0,71%
Malayalam	0,40%	0,53%
Bengalí	0,36%	0,48%
Telugu	0,29%	0,40%
Gujarati	0,27%	0,37%
Kannada	0,24%	0,34%
Urdu	0,14%	0,20%
Panjabi	0,13%	0,19%
Oriya	0,07%	0,09%
Sinhala	0,05%	0,09%

CONCLUSIONES

La muestra de dominios explorados es demasiado pequeña para sacar conclusiones definitivas; sin embargo, parece que está surgiendo un patrón, y cualquier contraejemplo será bienvenido, pero el riesgo es que sea más la excepción que la regla.

Ante datos tan desalentadores sobre las lenguas indígenas y, en general, sobre las lenguas locales, uno se pregunta legítimamente por qué, si las organizaciones internacionales y las organizaciones de la sociedad civil han priorizado el fomento de los contenidos locales desde los inicios de la web, durante tres décadas, los resultados son tan insignificantes.

Es una oportunidad para recordar que proponer políticas públicas sin indicadores capaces de diagnosticar el estado inicial y medir el progreso no es una manera apropiada de gestionar y es exponerse a sorpresas violentas.

Podemos concluir recordando las recomendaciones para políticas definidas, en su mayoría, hace 35 años y que siguen siendo válidas y muy necesarias:

- Sensibilizar a las partes interesadas del ecosistema web para animarlas a añadir versiones en idiomas locales a sus sitios web (en un momento en que los costes del multilingüismo web se han reducido drásticamente), no como una medida de solidaridad, sino como una opción comercial eficaz para ampliar su alcance de mercado y sus ingresos.

- Capacitar a los ciudadanos en alfabetización mediática e informacional, con un fuerte componente sobre el "por qué" y el "cómo" crear contenido en lenguas nativas.
- Otorgar premios a concursos y programas de creación de contenidos locales, así como a todas las iniciativas creativas posibles destinadas a fomentar la creación de contenidos locales.
- Ejemplos de iniciativas creativas: sensibilizar y capacitar a los agentes culturales para que incorporen sistemáticamente un componente virtual en idioma local en sus actividades...
- Y probablemente se requiera un estudio serio para comprender y analizar las razones del fracaso en esa prioridad máxima y definir medidas más efectivas para mejorar y alcanzar el objetivo identificado y acordado.

REFERENCIAS:

- Páginas de MECILDI que incluyen acceso a todos los resultados publicados: <https://obdilci.org/proyectos/mecildi/>
- Pimienta D., Field M., 2026, *MECILDI: A Multilingual-Aware Computational Framework for Measuring the Language Distribution of Web Content*, en fase de revisión por pares en Plos One. La versión en francés está disponible en: <https://www.obdilci.org/wp-content/uploads/2026/08/MECILDI-V1-FR.pdf>
- Estudios sobre el multilinguismo de la web. <https://www.obdilci.org/proyectos/otros/mlreports-4/>

ANEXO: Resultados completos para Cabo Verde (.cv)

INDICADORES	Definición del indicador
% SITIOS (L)	Sitios web que tienen una versión lingüística en lengua L
%PÁGINAS (L)	Porcentaje de páginas web en L (número de sitios en L dividido por el número de versiones lingüísticas)
MultiR	Tasa de multilingüismo (número de versiones lingüísticas dividido por el número de sitios web)
%Multi	Porcentaje de sitios web multilingües
AvgL	Número medio de idiomas por sitio web multilingüe
ERR-TOT	Porcentaje total de sitios web no procesados
ERR-DNS	Dominio no accesible
ERR-BLK	Sitios válidos, pero no permiten el acceso por programa.
ERR-CFG	Estacionado o en construcción
ERR-HTTP	Error del sitio web
ERR-LANG	Sitio web válido, pero falla la detección de lengua.
MonoL	Porcentaje de sitios web monolingües
GooT	Porcentaje de sitios web que utilizan el widget de Google Translate
HDEF	Factor de extrapolación derivado heurísticamente.
Lang=	Porcentaje de sitios web que utilizan el atributo HTML Lang=
HrefLang =	Porcentaje de sitios web que utilizan el atributo HTML HrefLang =
IDN	Porcentaje de sitios con nombres de dominio internacionalizados
NON-ID	Idioma no identificado por el algoritmo de detección

Número de sitios analizados: **2 429**

Número de sitios validados para el proceso: **1 212**

MultiR	1.79
%Multi	23,98%
Promedio L	4.30
MonoL	76,02%
GooT	1,97%
HDEF	2.62
Idioma=	79,51%
Lang=Error	11,39%
IDN	0,00%
ERR-DNS	18,37%
ERR-BLK	2,87%
ERR-CFG	1,76%
ERR-HTTP	6,03%
ERR-LANG	53,10%
ERR-TOT	82,13%

DISTRIBUCIÓN DE LENGUAS	%PÁGINAS	%SITIOS
TOTAL	100,00%	178,48%
Inglés	33,16%	59,18%
Portugués	25,12%	44,83%
Francés	5,86%	10,45%
Indonesio	4,44%	7,93%
Vietnamita	3,91%	6,97%
Español	2,10%	3,76%
Alemán	1,80%	3,21%
Chino	1,68%	3,01%
Holandés	1,19%	2,12%
Japonés	1,15%	2,05%
Italiano	1,03%	1,84%
Ruso	0,96%	1,71%
Árabe	0,84%	1,50%
Coreano	0,80%	1,43%
Turco	0,73%	1,30%
Polaco	0,69%	1,23%
Tailandés	0,61%	1,09%
Búlgaro	0,50%	0,89%
Checo	0,46%	0,82%
Malayo	0,46%	0,82%
Rumano	0,46%	0,82%
Sueco	0,46%	0,82%
Danés	0,34%	0,61%
Griego moderno	0,34%	0,61%
Finlandés	0,34%	0,61%
Ucranio	0,34%	0,61%
Latín	0,27%	0,48%
Persa	0,27%	0,48%
Bengalí	0,23%	0,41%
Estonio	0,23%	0,41%
Hindi	0,23%	0,41%
Húngaro	0,23%	0,41%
Javanés	0,23%	0,41%
Letón	0,23%	0,41%
Lituano	0,23%	0,41%
Mongol	0,23%	0,41%
Birmanio	0,23%	0,41%
Noruego	0,23%	0,41%

Serbio	0,23%	0,41%
Tagalo	0,23%	0,41%
Non-ID	0,19%	0,34%
Vasco	0,15%	0,27%
Irlandés	0,15%	0,27%
54 lenguas mas	0,11%	0,20%

Notas:

- El código ISO639-3 para Kabuverdianu (kea) no está presente en los campos HTML Lang o Hreflang.
- El código de país (cv) puede haberse utilizado como código de lengua en raras ocasiones, como en <https://ipc.cv/cv/>.
- Los sitios web en indonesio aparecen en redirecciones, principalmente sitios de juegos.
- El archivo .cv se vende en todo el mundo para que las personas almacenen su currículum profesional con el formato JohnSmith.cv. El administrador de esos sitios los ha marcado como no aptos para la navegación de robots y, por lo tanto, no ha afectado los resultados, excepto por un enorme porcentaje de ERR-LANG, ya que a MECILDI le fue imposible procesar su contenido.
- Esos sitios representan más de la mitad del total de sitios web.