

Documents complémentaires pour :
MECILDI : une plateforme d'observation multilingue pour
mesurer la distribution linguistique du contenu Web

Daniel Pimienta and Mike Field

Daniel Pimienta

Observatory of Linguistic and Cultural Diversity on the Internet (OBDILCI), Nice, France

pimienta@funredes.org

<https://orcid.org/0000-0001-7314-6471>

Mike Field

Mike Field Enterprises, Toronto, Canada

mike@mikefield.ca

<https://orcid.org/0009-0009-5892-7937>

Table des matières

S 1. Langues prises en charge par le détecteur de langue de MachineTranslation.com (Tomedes) et leurs codes ISO 639-1 et ISO 639-3 correspondants	3
S 2. Expressions de sites en construction.....	9
S 3. Expressions clés pour fournisseurs d'hébergement web	10

S 1. Langues prises en charge par le détecteur de langue de MachineTranslation.com (Tomedes) et leurs codes ISO 639-1 et ISO 639-3 correspondants

Cette section répertorie toutes les langues prises en charge par le détecteur de langue MachineTranslation.com (Tomedes) utilisé dans la plateforme MECILDI, ainsi que leurs identifiants ISO 639-1 et ISO 639-3 correspondants. Ces tableaux fournissent l'ensemble des références sous-jacentes à toutes les opérations de détection de langue décrites dans le texte principal.

Note : Dans la colonne « *Entrée Ethnologue correspondante* », les entrées à trois lettres correspondent aux codes ISO 639-3 désignant les macrolangues. Les cellules vides indiquent qu'aucune entrée Ethnologue correspondante n'a été trouvée.

Tableau 1 Langues prises en charge par MachineTranslation.com (Tomedes)

Code ISO 639-3	Code ISO 639-1	Nom de la langue (MachineTranslation.com)	Entrée Ethnologue correspondante (en anglais)
aar	aa	Afar	
abk	ab	Abkhaze	Abkhaz
afr	af	Afrikaans	
aka	ak	Akan	aka
amh	am	Amharique	
aqg	io	Ido	Dialect of Arigdi
ara	ar	Arabe	ara
arg	an	Aragonais	
arz	egy	Égyptien	Arabic, Egyptian
asm	as	Assamais	
ast		Asturien	
ava	av	Avaric	Avar
aym	ay	Aymara	aym
aze	az	Azerbaïdjanais	aze
bak	ba	Bachkir	Bashkort
bar		Bavarois	
bel	be	Biélorusse	
ben	bn	Bengali	
bik		Bikol	bik
bis	bi	Bislama	
bod	bo	Tibétain	Tibetan, Central
bos	bs	Bosniaque	
bpy		Bishnupriya	Bishnupuriya
bre	br	Breton	
bua		Bouriate	bua
bug		Bugis	Bugis

bul	bg	Bulgare	
cat	ca	Catalan	
cbk		Chavacano	
ceb		Cebuano	
ces	cs	Tchèque	
che	ce	Tchéchène	
chr		Cherokee	
chv	cv	Tchouvash	
cor	kw	Cornouaillais	
cos	co	Corse	
crs		Seselwa créole français	Seychelles French Creole
cym	cy	Gallois	
dan	da	Danois	
deu	de	Allemand	German, Standard
div	dv	Dhivehi	Maldivian
dsb		Bas-Sorabe	Sorbian, Lower
dzo	dz	Dzongkha	
egl		Émilien	
ell	el	Grec moderne	Greek
eng	en	Anglais	
epo	eo	Espéranto	
est	et	Estonien	est
eus	eu	Basque	
fao	fo	Féroé	
fas	fa	Persan	fas
fij	fj	Fidjien	
fin	fi	Finlandais	
fra	fr	Français	
frf		Frison du nord	Frisian, Northern
fry	fy	Frison occidental	Frisian
gla	gd	Gaélique écossais	
gle	ga	Irlandais	
glg	gl	Galicien	
glv	gv	Manx	
grn	gn	Guarani	grn
guj	gu	Gujarati	
hat	ht	Haïtien	Haitian Creole
hau	ha	Haoussa	
haw		Hawaïen	
hbs	sh	Serbo- croate	hbs
heb	he	Hébreu	
hif	hif	Fidji hindi	
hin	hi	Hindi	
hmn		Hmong	hmn
hob		Mari	

hrv	hr	Croate	
hsb		Haut-sorabe	Sorbian, Upper
hun	hu	Hongrois	
hye	hy	Arménien	
ibo	ig	Igbo	
iku	iu	Inuktitut	iku
ila	ia	Interlingua	Not referenced
ile	ie	Interlingue	Not referenced
ilo		Iloko	Ilocano
ind	id	Indonésien	
ipk	ik	Inupiaq	ipk
isl	is	Islandais	
ita	it	Italien	
jav	iv	Javanais	
jpn	ja	Japonais	
kal	kl	Kalaallisut	Greenlandic
kan	kn	Kannada	
kas	ks	Cachemiri	
kat	ka	Géorgien	
kaz	kk	Kazakh	
kha		Khasi	
khm	km	Khmer	
kin	rw	Kinyarwanda	
kok		Konkani	kok
kom	kv	Komi	kom
kor	ko	Coréen	
krc		Karachay balkar	Karachay-Balkar
kur	ku	Kurde	kur
kyr	ky	Kirghiz	Kyrgyz
lah		Lahnda	lah
lao	lo	Lao	
lat	la	Latin	
lav	lv	Letton	lav
lez		Lezghien	Lezgi
lif		Limbu	
lim	li	Limbourgeois	Limburgish
lin	ln	Lingala	
lit	lt	Lituanien	
lmo		Lombard	
lrc		Luri du nord	Luri, Northern
ltz	lb	Luxembourgeois	
lug	lg	Ganda	
mai		Maïthili	
mal	ml	Malavalam	
mar	mr	Marathi	

mfe		Morisyen	
mkd	mk	Macédonien	
mlg	mg	Malgache	mlg
mlt	mt	Maltais	
mon	mn	Mongol	mon
mri	mi	Maori	
msa	ms	Malais	msa
mwł		Mirandais	
mya	my	Birman	
myv		Erzya	
mzn		Mazanderani	Mazandarani
nap		Napolitain	Napoletano
nau	na	Nauru	Nauruan
nbl	nr	Ndébéle du sud	Ndebele
nds		Bas allemand	Saxon, Low
new		Népal bhasa	Newar
nhe		Nahuatl Huasteca oriental	Nahuatl, Eastern Huasteca
nld	nl	Néerlandais	
nor	no	Norvégien	nor
npi	ne	Népalais	
nso		Pédi	Sotho, Northern
nya	ny	Nyanja	Chichewa
oci	oc	Occitan	
ori	or	Oriya	ori
orm	om	Oromo	orm
oss	os	Ossète	Ossetic
ɒam		Pampang	Kapampangan
ɒan	pa	Panjab	Punjab, Eastern
pfl		Pfaelzisch	Palatinate Franconian
pms		Piémontais	Piedmontese
pol	pl	Polonais	
por	pt	Portugais	
pus	ps	Pachtoune	pus
que	qu	Quechua	que
roh	rm	Romansh	
ron	ro	Roumain	
rue		Ruthène	
run	rn	Rundi	
rus	ru	Russe	
sag	sg	Sango	
sah		Yakoute	
san	sa	Sanskrit	san
scn		Sicilien	
sco		Écossais	
sin	si	Cinghalais	

slk	sk	Slovaque	
slv	sl	Slovène	Slovene
smo	sm	Samoan	
sna	sn	Shona	
snd	sd	Sindhi	
som	so	Somali	
sot	st	Sotho du sud	Sotho, Southern
spa	es	Espagnol	
sqi	sq	Albanais	sqi
srd	sc	Sarde	srd
srp	sr	Serbe	
ssw	ss	Swati	
sun	su	Soundanais	Sunda
swe	sv	Suédois	
swh	sw	Swahili	
syc		Syriaque	syr
tam	ta	Tamoul	
tat	tt	Tatar	
tel	te	Telugu	
tgk	tg	Tadjik	
tgl	tl	Tagalog	
tha	th	Thaïlandais	
tir	ti	Tigrinya	Tigrigna
tlh		Klingon	Not referenced
tog	to	Tonga	
tsn	tn	Tswana	Setswana
tso	ts	Tsonga	
tuk	tk	Turkmènes	
tur	tr	Turc	
tyv		Touvien	Tuvan
uig	ug	Ouïghour	Uyghur
ukr	uk	Ukrainien	
urd	ur	Ourdou	
uzb	uz	Ouzbek	uzb
vec		Vénitien	
ven	ve	Venda	
vep		Veps	
vie	vi	Vietnamien	
vls		Flamand	West Flemish
vol	vo	Volapük	Not referenced
war		Waray	Waray-Waray
wln	wa	Wallon	
wol	wo	Wolof	
xal		Kalmouk	Kalmyk-Oirat
xho	xh	Xhosa	

xmf		Mingrélien	
vid	yi	Yiddish	vid
vor	vo	Yoruba	
zha	za	Zhuang	zha
zho	zh	Chinois	zho
zul	zu	Zoulou	
zza		Zaza	zza
	got	Gothique	Non référencé
	jbo	Lojban	Non référencé
	nn	Norvégien nynorsk	Non référencé

S 2. Expressions de sites en construction

Ces expressions servent à la détection des erreurs, comme décrit dans la section 3.4 relative à la gestion des erreurs. Pour figurer dans cette liste, chaque expression doit apparaître plus de 300 000 fois dans les résultats de recherche Google.

1. Site under construction
2. Web site under construction
3. Page under construction
4. Site under maintenance
5. Web site under maintenance
6. Site coming soon
7. Web site coming soon
8. Site en construction
9. Site en maintenance
10. Site web en maintenance
11. Sitio en mantenimiento
12. Sitio en construcción
13. webseite im Aufbau
14. Сайт в разработке
15. 网站建设中
16. قيد الإنشاء
17. 準備中

S 3. Expressions clés pour fournisseurs d'hébergement web

Ces phrases sont utilisées pour la détection des erreurs, comme décrit dans la section 3.4 sur la gestion des erreurs.

1. GoDaddy
2. IONOS
3. Bluehost
4. HostGator
5. DreamHost
6. Newfold Digital
7. Amazon Web Services
8. Google Cloud
9. Microsoft Azure
10. OVHcloud
11. Scaleway
12. Hetzner
13. Infomaniak
14. Wix
15. Squarespace
16. Webflow
17. Strato
18. Netcup
19. All-Inkl
20. 阿里云
21. 阿里云计算
22. 腾讯云
23. 华为云
24. 新网
25. 中国电信
26. 中国联通
27. エックスサーバー
28. さくらインターネット
29. ロリポップ!
30. Conoha
31. お名前.com
32. インターネット
33. コミュニケーションズ