

Observatori de la Diversitat Cultural

Wikipedia Cultural Diversity Observatory (WCDO)

[<https://meta.wikimedia.org/wiki/WCDO>]

Marc Miquel {marcmiquel@gmail.com}

Usuari:marcmiquel

Viquitrobada, Manresa, 2018



El problema

Viquipèdia (projecte) no reflecteix prou la diversitat cultural del món.



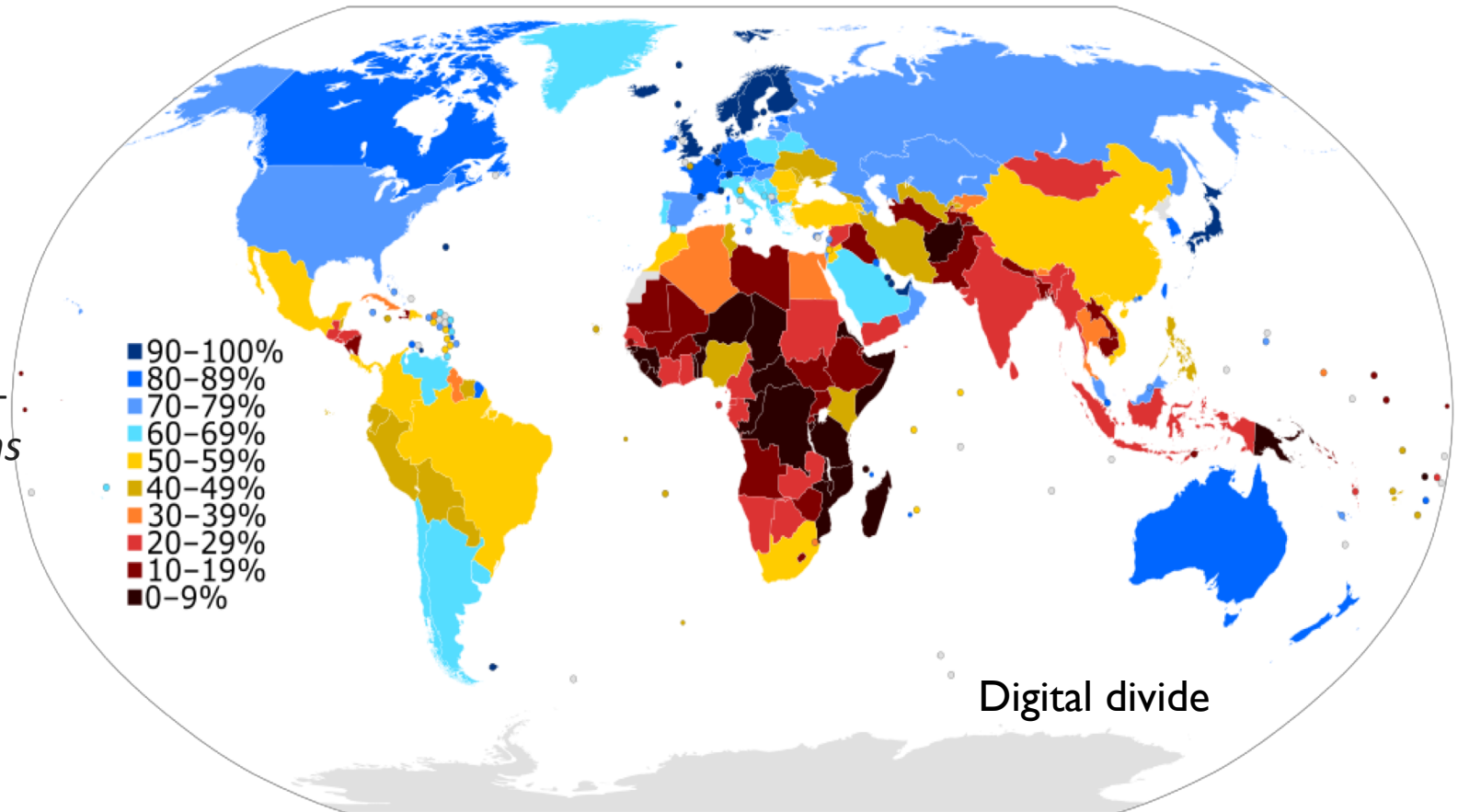
Algunes veus no hi són o no estan representades



1. Molts article que haurien de descriure la diversitat del món no existeixen perquè no totes les llengües tenen Viquipèdia o la gent no hi pot contribuir.

Sabem que això es deu a factors com l'escletxa digital, la reputació de la llengua, estatus legal, entre d'altres.

Van Dijk, Z. (2009). Wikipedia and lesser-resourced languages. *Language Problems and Language Planning*, 33(3), 234-250.



2. Hi ha molts articles que representen les peculiaritats culturals i tracten allò específic de molts contextos del món... Però ningú fora d'aquestes Viquipèdies pot o sap accedir-hi.

Aquí és on realment podem treballar.

Solució proposada

Wikipedia Cultural Diversity Observatory (WCDO).

“un espai en comú per **investigadors i activists** per estudiar i **lluitar contra els forats de contingut** i promoure la diversitat cultural”.

- **Reconeixement dels problemes (Awareness)**
- **Eines per treballar-hi (Solutions)**



<http://wcdo.wmflabs.org> (beta 1)

És impossible tapar tots els forats de contingut entre llengües.



Objectiu de l'Observatori

“Cada Viquipèdia asseguri un mínim de cobertura del contingut relatiu al context geogràfic/cultural de cada altra llengua”.



Proposo que cada Viquipèdia tingui 100 articles sobre el context cultural de cada altra llengua.

Això són uns 28-30 mil articles.



Top CCC articles lists

De cada llengua, aquells articles del seu context cultural els quals són més rellevants d'acord amb una sèrie de característiques:

- List = [editors, featured, geolocated, keywords, women, men, created_first_three_years, created_last_year, pageviews, discussions]
- Country_origin (optional) = ISO3166 code
- Lang_origin = wikicode
- Lang_target = wikicode

http://wcdо.wmflabs.org/top_ccc_articles/

http://wcdо.wmflabs.org/top_ccc_articles/?list=men&lang_origin=pl&lang_target=uk

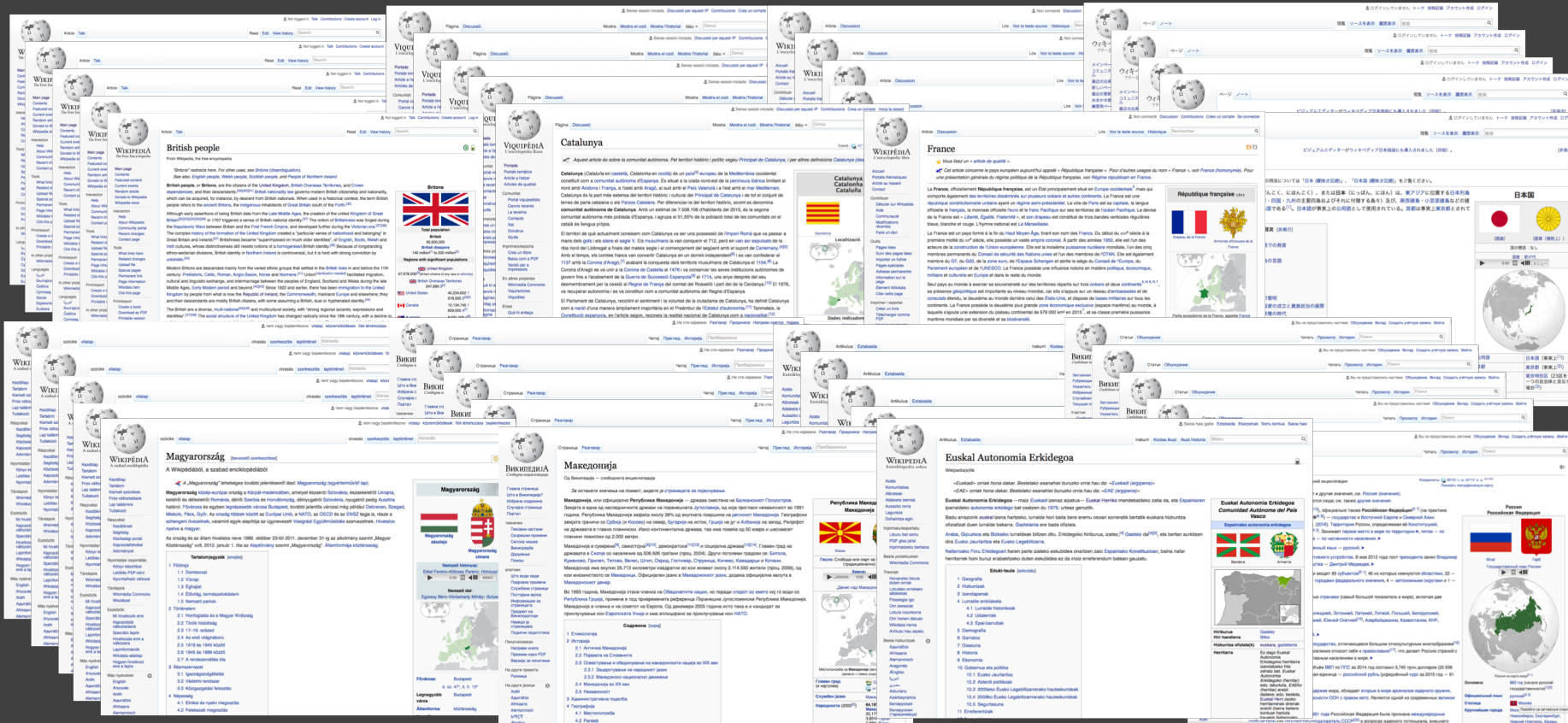
Top articles in Romanian CCC by editors and their availability in Polish

In this page, you can consult Top CCC articles list from any country or language CCC generated with the latest CCC dataset. The following table shows the Top 500 articles from Romanian CCC by editors and their availability in Polish Wikipedia. They are sorted by the feature **editors**. The rest of columns present complementary features that are explicative of the article relevance. In particular, number of Inlinks from CCC highlights the article importance in terms of how much is required by other articles from the Cultural Context Content.

The available Top CCC articles lists are: list of CCC articles with most number of editors (**Editors**), list of CCC articles with featured article distinction (**Featured**), most bytes and references (weights: 0.8, 0.1 and 0.1 respectively), list of CCC articles with geolocation with most links coming from CCC, list of CCC articles with keywords on title with most bytes (**Bytes**), list of CCC articles categorized in Wikidata as women with most edits (**Women**), list of CCC articles categorized in Wikidata as men with most edits (**Men**), list of CCC articles created during the first three years and with most edits (**First 3Y.**), list of CCC articles created during the last year and with most edits (**Last Y.**), list of CCC articles with most pageviews during the last month (**Pageviews**), list of CCC articles with most edits in talk pages (**Discussions**). The table's last column shows the title the article has in its target language, in blue when it exists and in red as a proposal generated with the Wikimedia Content Translation tool or as an existing Wikidata label in the same language.

It is possible to query any list by changing the URL parameters. You need to specify the list parameter (editors, featured, geolocated, keywords, women, men, created_first_three_years, created_last_year, pageviews and discussions), the language target parameter (as lang_target and the language wikicode), the language origin (as lang_origin and the language wikicode), and, optionally to limit the scope of the selection, the country_origin parameter as part of the CCC (as country_origin and the country ISO3166 code). In case no country is selected, the default is 'all'.

N°	Romanian Title	Editors	Pageviews	Bytes	References	Wikidata Properties	Interwiki Links	Inlinks from CCC	Creation Date	Other Languages	Polish Title
1	România	1064	4774	213456	410	137	298	42303	2003-07-15	en , fr , de , ru	Rumunia
2	București	802	2111	167239	193	63	205	7117	2003-07-26	en , fr , de , ru	Bukareszt
3	Republica Moldova	724	1503	185309	187	123	278	4958	2003-08-02	en , fr , de , ru	Mołdawia
4	FC Dinamo București	712	417	70889	62	14	41	482	2005-05-05	en , fr , de , ru	FC Dinamo Bukareszt
5	Iași	613	739	110146	74	39	94	2043	2004-01-31	en , fr , de , ru	Jassy
6	Cluj-Napoca	606	1063	115794	107	52	99	1289	2003-11-11	en , fr , de , ru	Kluż-Napoka
7	FC Rapid București	568	282	83241	37	12	33	309	2004-06-05	en , fr , de , ru	FC Rapid Bukareszt
8	Brașov	547	757	93804	62	41	92	1490	2004-01-31	en , fr , de , ru	Braszów
9	Timișoara	519	675	205305	182	46	94	1410	2004-02-02	en , fr , de , ru	Timișoara



Com ho fem això?
Seleccióem els **Continguts del Context Cultural (CCC)**, és a dir, els articles relacionats amb els contextos culturals dels editors en cada edició lingüística (tradicions, llenguatge, política, agricultura, biografies, llocs, esdeveniments, etc.).
Això significa associar cada idioma als territoris on es parla oficialment o on és natiu, i després, seleccionar articles que es relacionen amb cada territori.

3. Metodologia

ccc_setup_language_context_mapping.py

Per fer CCC cal (i) una bbdd de **Language-Territories Mapping** i (ii) extreure continguts de les viquipèdies i etiquetar-los en funció de la *relació* amb CCC.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	territoryname	territorynameNative	QitemTerritory	languagenam	Wiki	demon	demon	ISO3166	ISO31662	region	country	ind	lan	official	nt
1	Afar	Afar	Q193494	Afar	aa			ET	ET-AF	yes	Ethiopia	yes	2	regional	0
2	Somali	Somali	Q202800	Afar	aa			ET	ET-SO	yes	Ethiopia	yes	2	regional	0
3	Amhara	Amhara	Q203009	Afar	aa			ET	ET-AM	yes	Ethiopia	yes	2	regional	0
4	Ali Sabieh	Ali Sabieh	Q821008	Afar	aa			DJ	DJ-AS	yes	Djibouti	yes	5	no	0
5	Arta	Arta	Q705941	Afar	aa			DJ	DJ-AR	yes	Djibouti	yes	5	no	0
6	Obock	Obock	Q844929	Afar	aa			DJ	DJ-OB	yes	Djibouti	yes	5	no	0
7	Dikhil	Dikhil	Q283979	Afar	aa			DJ	DJ-DI	yes	Djibouti	yes	5	no	0
8	Debubawi K'eyih	Debubawi K'eyih	Q27728	Afar	aa			ER	ER-DU	yes	Eritrea	yes	5	no	0
9	Semenawi K'eyi B	Semenawi K'eyi Bahri	Q27910	Afar	aa			ER	ER-SK	yes	Eritrea	yes			
10	Abkhazia	Аԥсны	Q23334	Abkhaz	ab	Abkhaz		GE	GE-AB	yes	Georgia	yes	2	regional	1
11	Aceh	Acèh	Q1823	Aceh	ace			ID	ID-AC	yes	Indonesia	yes	6	no	0
12	Sumatera Utara	Sumatra Baròh	Q2140	Aceh	ace			ID	ID-SU	yes	Indonesia	yes	6	no	0
13	Republic of Adyghe	Адыгæ	Q3734	Adyghe	ady			RU	RU-AD	yes	Russian Federation	yes	2	regional	1
14	Krasnodar Krai	Краснодар край	Q3680	Adyghe	ady			RU	RU-KDA	yes	Russian Federation	yes	2	regional	1
15	Karachay-Cherkess	Къарачае-Черкес	Q5328	Adyghe	ady			RU	RU-KC	yes	Russian Federation	yes	2	regional	1
16	South Africa	Suid-Afrika	Q258	Afrikaans	af	South Afri	Suid-Afrika	ZA		no	South Africa	yes	1	national	1
17	Central	Sentraal distrik	Q57525	Afrikaans	af			BW	BW-CE	yes	Botswana	yes	5	no	1
18	Ghanzi	Ghanzi	Q57571	Afrikaans	af			BW	BW-GH	yes	Botswana	yes	5	no	1
19	Kgalagadi	Kgalagadi	Q57581	Afrikaans	af			BW	BW-KG	yes	Botswana	yes	5	no	1
20	Kgatleng	Kgatleng	Q57593	Afrikaans	af			BW	BW-KL	yes	Botswana	yes	5	no	1
21	Southern	Suid distrik	Q57609	Afrikaans	af			BW	BW-SO	yes	Botswana	yes	5	no	1
22	Botswana	Botswana	Q963	Afrikaans	af	Motswana;	Botswana	BW		no	Botswana	yes	5	no	1
23	Ghana	Ghana	Q117	Akan	ak	Ghanaian		GH		no	Ghana	yes	3	no	1
24	Switzerland	Schweiz	Q39	German, Swiss	als	Swiss		CH		no	Switzerland	yes	5	no	0
25	Vorarlberg	Vorarlberg	Q38981	German, Swiss	als			AT	AT-8	yes	Austria	yes	5	no	0
26	Champagne-Arde	Champagne-Ardenne	Q14103	German, Swiss	als			FR	FR-G	yes	France	yes	6	no	0
27	Lorraine	Lothringen	Q1137	German, Swiss	als			FR	FR-M	yes	France	yes	6	no	0
28	Alsace	Elsass	Q1142	German, Swiss	als			FR	FR-A	yes	France	yes	6	no	0
29	Baden-Württemb	Baden-Württemberg	Q985	German, Swiss	als			DE	DE-BW	yes	Germany	yes	5	no	0

Language Territories mapping
fulla de càlcul amb 1783 files.

(i) Wikidata Language Qitem, Language name, Language name in Native language, the ISO 639 code, the associated territories at country level (ISO 3166 code, English name, Native language name, demonym, Qitem) or at first subdivision (ISO 3166-2 code, English name, Native language name, demonym, Qitem) according to the information generated by Ethnologue.

https://wcd0.wmflabs.org/language_territories_mapping/

Per exemple:

Italy	Italia	Q38	Italian	it	Italian	italiano;ita	IT		no	Italy
Istria	Istria	Q58268	Italian	it			HR	HR-18	yes	Croatia
San Marino	San Marino	Q238	Italian	it	Sammarinese	sammarin	SM		no	San Marino
Piran	Pirano	Q1382	Italian	it			SI	SI-090	yes	Slovenia
Izola	Isola	Q15877	Italian	it			SI	SI-040	yes	Slovenia
Graubünden	grigioni	Q11925	Italian	it			CH	CH-GR	yes	Switzerland
Ticino	ticino	Q12724	Italian	it			CH	CH-TI	yes	Switzerland
Vatican City	vaticano	Q237	Italian	it			VA		no	Vatican State
Sweden	Sverige	Q34	Swedish	sv	Swedish	svensk;sve	SE		no	Sweden
Åland s	Åland	Q5689	Swedish	sv	Ålandic	Ålänning	FI	FI-01	yes	Finland
Kymenlaakso	Kymmenedalen	Q5698	Swedish	sv			FI	FI-09	yes	Finland
Ostrobothnia	Österbotten	Q5702	Swedish	sv			FI	FI-12	yes	Finland

Territories where the language is spoken as **native or with official status**

(i) Wikidata Language Qitem, Language name, Language name in Native language, the ISO 639 code, the associated territories at country level (ISO 3166 code, English name, Native language name, demonym, Qitem) or at first subdivision (ISO 3166-2 code, English name, Native language name, demonym, Qitem) according to the information generated by Ethnologue.

https://wcd0.wmflabs.org/language_territories_mapping/

Algunes d'aquestes estratègies són fiables i d'altres et donen dades potencials.

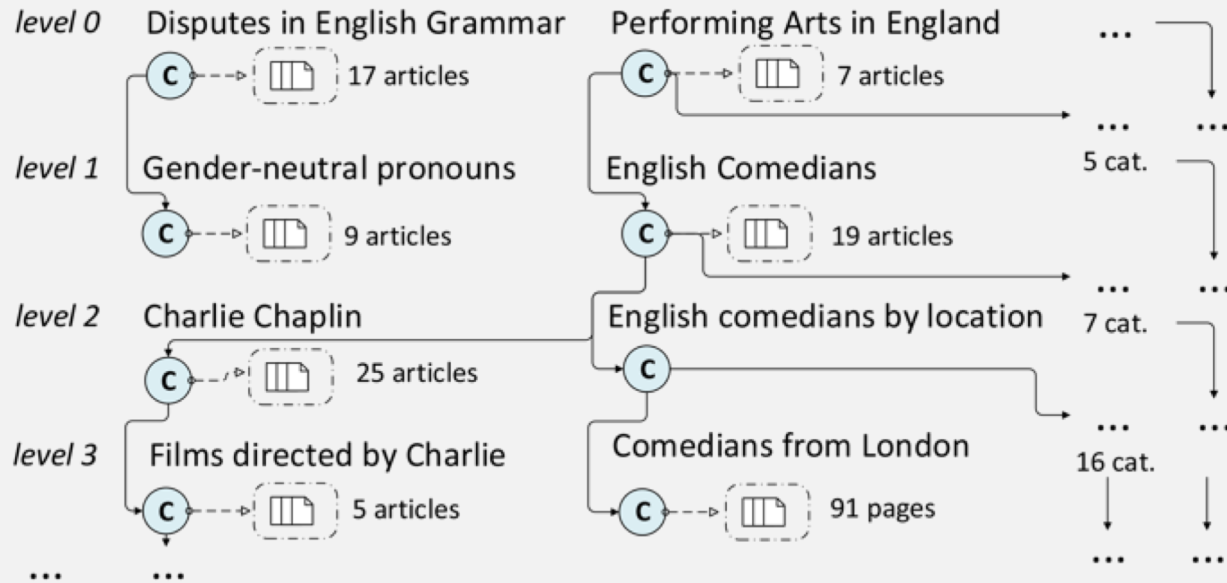
The screenshot shows the Wikipedia article for "English literature". The title "English literature" is prominently displayed at the top. Below the title, there is a brief introductory sentence: "From Wikipedia, the free encyclopedia". The main text of the article discusses the focus on English-language literature and mentions writers from Scotland, Wales, and Ireland. A table of contents is visible on the left side, listing various periods and sub-topics like "Old English literature", "Middle English literature", "English Renaissance", "Jacobean period", "Late Renaissance", "Restoration Age", and "18th century". There is also a grid of portraits of English-language writers, including Geoffrey Chaucer, William Shakespeare, Jane Austen, Mark Twain, Virginia Woolf, T. S. Eliot, Vladimir Nabokov, Toni Morrison, and Salman Rushdie.

The screenshot shows the Wikipedia article for "Times Square". The title "Times Square" is prominently displayed at the top. Below the title, there is a brief introductory sentence: "From Wikipedia, the free encyclopedia". The main text of the article describes Times Square as a major commercial intersection, tourist destination, and entertainment center in Midtown Manhattan. It mentions the annual New Year's Eve ball drop and the site's location at the intersection of Broadway and Seventh Avenue. A table of contents is visible on the left side, listing "History" and "Early history". There are also two photographs of Times Square: one from 2009 and one from 2013 showing Broadway show billboards.

- Paraules clau (gentilici/nom del territori) al títol és CCC fiable

- Geubicació en alguns dels territoris és CCC fiable

Keywords {English, England, Ireland, Irish, etc.}



Baixar per les categories amb paraules clau

Dylan Moran

From Wikipedia, the free encyclopedia

Dylan William Moran (/ˈmɔːrən/; born 3 November 1971)^[1] is an Irish comedian, writer, actor and filmmaker. He is best known for his observational comedy, the television sitcom *Black Books* (in which he starred and co-wrote) and his work with *Simon Pegg* in *Shaun of the Dead* and *Run Fatboy Run*. He appeared as one of the two lead characters in the Irish black comedy titled *A Film with Me in It* in 2008.

Moran's most recent film is *Calvary*, an Irish black comedy drama film written and directed by John Michael McDonagh. Moran is a regular performer at national and international comedy festivals including the Edinburgh Festival Fringe, Just for Laughs Montreal Comedy Festival, the Melbourne International Comedy Festival and the Kilkenny Comedy Festival. In 2007, Moran was voted the 17th greatest stand-up comic on Channel 4's 100 Greatest Stand-Ups and again in the updated 2010 list as the 14th greatest stand-up comic. He lives in Edinburgh with his wife, Elaine, and two children.

Contents [hide]

- Biography
 - 1.1 Early life
 - 1.2 Career
 - 1.3 Awards and commendations
- Filmography
 - 2.1 Film
 - 2.2 Television
- Stand-up DVDs
- References
- External links

Biography [edit]

Early life [edit]

Moran was born in Navan, County Meath, Ireland.^{[1][2][3][4]} He attended St. Patrick's Classical School, where he experimented early on with

- Ser en una subcategoria d'una categoria que conté una paraula clau és potencial CCC

Algunes propietats de Wikidata són CCC fiable

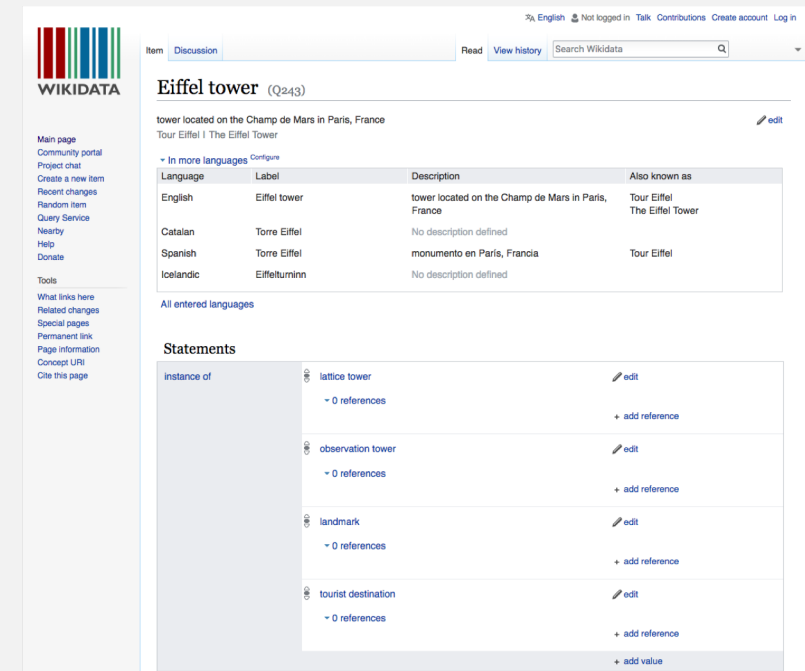
- **Location properties (location, located in administrative,...).**
- **Country properties (country of citizenship, of origin).**
- **Language properties (official language, native language...).**

Algunes propietats de Wikidata són potencial CCC

- **Affiliation properties (member of, educated at, employer,...).**
- **Has part (contains administrative entity, has part).**
- **Language properties (language of work, language used,...).**

Vincles:

- **El nombre i el percentatge d'enllaços / enllaços externs (enllaços entrants / sortints) a CCC és molt explicatiu de com es necessita un article per expandir CCC o dedicar-se a CCC.**



Wikidata page for 'Eiffel tower' (Q243). The page displays the item's description in French and English, and a table of labels in various languages. The 'Statements' section shows properties like 'instance of', 'observation tower', 'landmark', and 'tourist destination' with their respective values and reference counts.

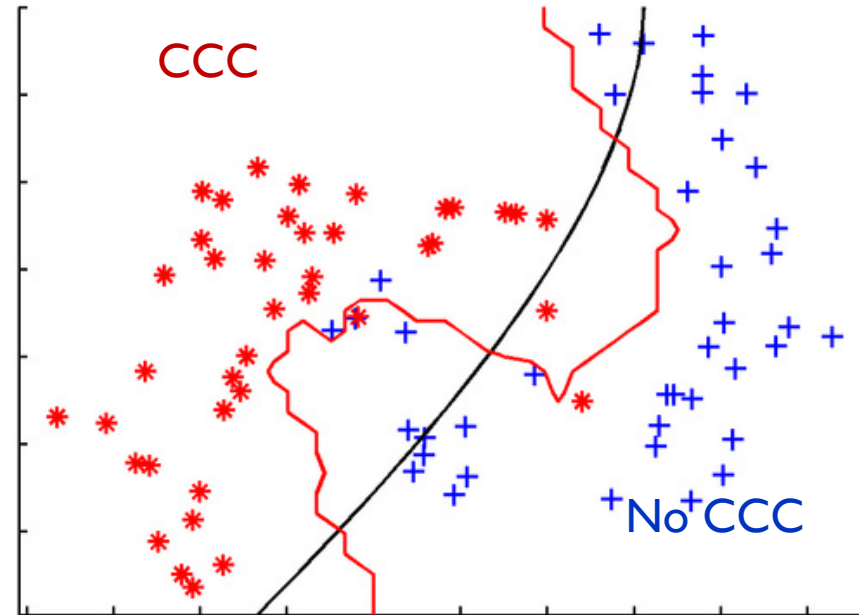
Language	Label	Description	Also known as
English	Eiffel tower	tower located on the Champ de Mars in Paris, France	Tour Eiffel The Eiffel Tower
Catalan	Torre Eiffel	No description defined	
Spanish	Torre Eiffel	monumento en París, Francia	Tour Eiffel
Icelandic	Eiffeltúminn	No description defined	

Property	Value	References	Action
instance of	lattice tower	0 references	edit / + add reference
	observation tower	0 references	edit / + add reference
	landmark	0 references	edit / + add reference
	tourist destination	0 references	edit / + add reference

Amb les característiques, establím un groundtruth (articles que sabem que ja són CCC i articles que sabem que no són CCC).

MACHINE LEARNING CLASSIFIER: RANDOM FOREST

- Training Data
- Testing Data



Els articles positius del classificador i la base CCC inicial constitueixen el CCC final. Es va realitzar una avaluació manual (cega) per determinar la qualitat de la selecció i els resultats eren en mitjana un 5% falsos positius i un 5% negatius falsos.

Project's Technical Overview

- **Wikimedia Cloud Server at Toolforge**

Servidor: <https://tools.wmflabs.org/admin/tool/wcdo>*Phabricator:* <https://phabricator.wikimedia.org/T193984>

Execució: crontab (cron job in shell) to execute the scripts on a **monthly basis**.

Scripts de Python :

ccc_setup_language_context_mapping.py (crea la base de dades d'equivalències territori-llengua).

ccc_selection.py (crea la base de dades principal ccc_current.db i els datasets).

stats_generation.py (crea la base de dades wcdo_data.db amb les estadístiques).

meta_update.py (actualitza amb ***pywikibot*** el portal de meta).

- **Datasets**

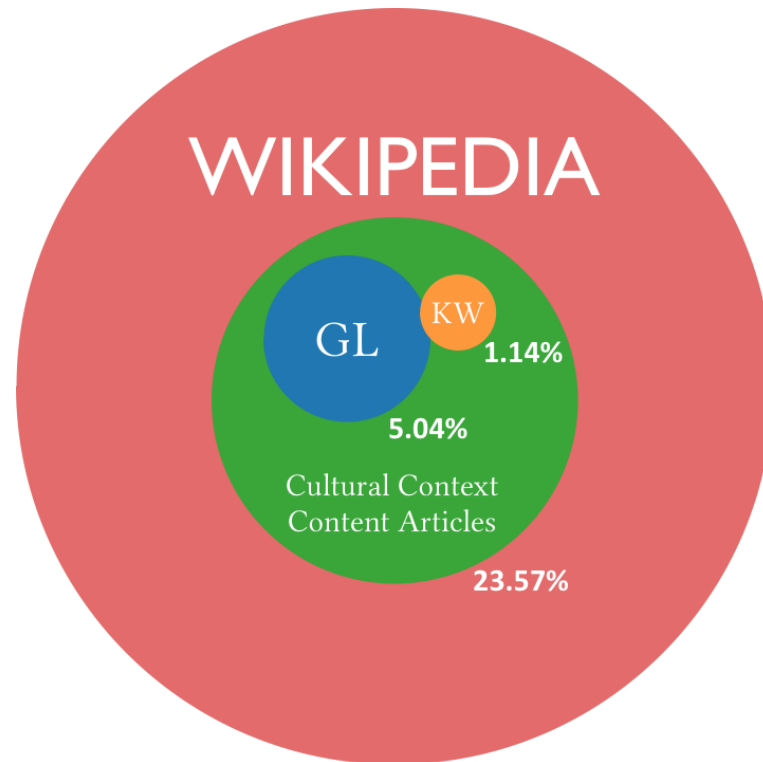
Disponibles a wcdo.wmflabs.org/datasets

- **Codi a Github**

Tot el codi, presentacions i papes són a: github.com/marcmiquel/WCDO

Tenint en compte les llengües principals, CCC és aproximadament una quarta part de la Viquipèdia.

El percentatge d'articles CCC és molt variable depenent de la llengua.



Quina és l'extensió o proporció del CCC a cada VP?

Llista de Viquipèdies per contingut de context cultural

http://wcdo.wmflabs.org/list_of_wikipedias_by_cultural_context_content

http://wcdo.wmflabs.org/list_of_language_territories_by_cultural_context_content

Lists of Wikipedias by Cultural Context Content

This page contains a list of all the current Wikipedia language editions ordered by their number of articles from their Cultural Context Content dataset that relate to territories where the language is spoken as official or as indigeneous.

For each language edition, statistics account for the number of articles of different CCC segments and their percentage computed in relation to the overall total number of Wikipedia articles. This is **(CCC art.)** and **CCC (%)** as the number of CCC articles and percentage, **CCC GL (%)** as the number of articles from CCC that are geolocated, **KW Title (%)** as the number of articles from CCC that contain specific keywords (language name, territory name or demonym) in their titles. Finally, **Region** (continent) and **Subregion** are introduced in order to contextualize the results.

<input type="checkbox"/>	N°	Language	Wiki	Articles	CCC art.	CCC %	Subregion	Region
<input type="checkbox"/>	295	Afar	aa	1	0	0	Sub-Saharan /	Africa
<input type="checkbox"/>	219	Abkhaz	ab	3471	143	4.12	Western Asia	Asia
<input type="checkbox"/>	91	Acehnese	ace	7478	4689	62.7	South-eastern	Asia
<input type="checkbox"/>	264	Adyghe	ady	544	32	5.88	Eastern Europe	Europe
<input type="checkbox"/>	66	Afrikaans	af	52138	12510	23.99	Sub-Saharan /	Africa
<input type="checkbox"/>	234	Akan language	ak	617	92	14.91	Sub-Saharan /	Africa
<input type="checkbox"/>	74	Alemannic	als	24820	10420	41.98	Western Europe	Europe
<input type="checkbox"/>	178	Amharic	am	14815	395	2.67	Sub-Saharan /	Africa
<input type="checkbox"/>	90	Aragonese	an	33081	4845	14.65	Southern Europe	Europe
<input type="checkbox"/>	197	Old English	ang	3058	262	8.57	Northern Europe	Europe
<input type="checkbox"/>	10	Arabic	ar	588607	240034	40.78	Western Asia	Asia

List of Language Territories by Cultural Context Content

This page contains each Wikipedia language edition Cultural Context Content divided in its territories according to the language territories mapping. Articles are assigned to territories according to the different strategies that have been used to include them into CCC. The label Not Assigned is for those articles which were not possible to classify.

For each territory, statistics account for the number of articles of different CCC segments and their percentage computed in relation to the overall total number of Wikipedia articles. This is **(CCC art.)** and **CCC (%)** as the number of CCC articles and percentage, **CCC GL (%)** as the number of articles from CCC that are geolocated, **KW Title (%)** as the number of articles from CCC that contain specific keywords (language name, territory name or demonym) in their titles. Finally, **Region** (continent) and **Subregion** are introduced in order to contextualize the results.

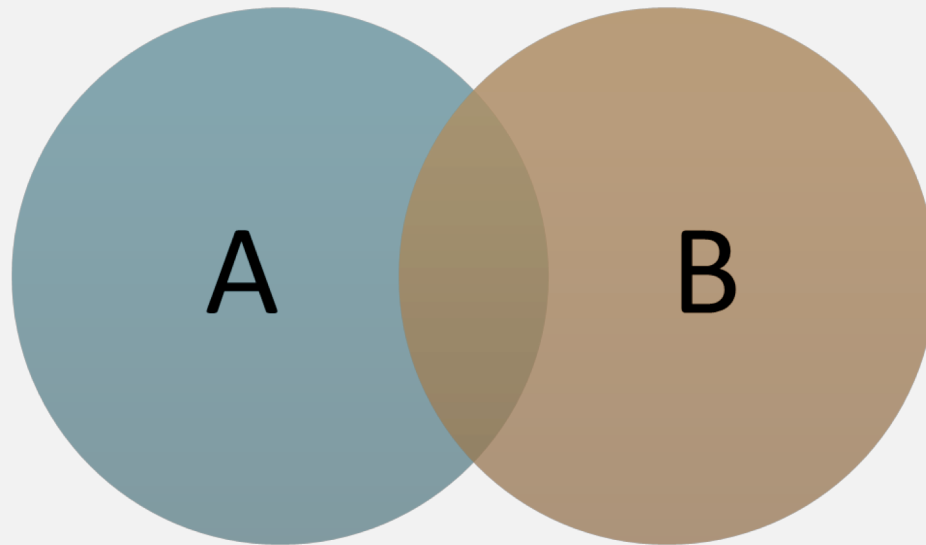
											FILTER ROWS
<input type="checkbox"/>	Qitem	Territory na	Language	Wiki	CCC art.	CCC %	CCC GL art	CCC art. K	ISO3166	ISO3166-2	subre
<input type="checkbox"/>	Q23334	Abkhazia	Abkhaz	ab	81	56.64	42	8	GE	GE-AB	Weste
<input type="checkbox"/>	Not Assigr		Abkhaz	ab	62	43.36	54	7			
<input type="checkbox"/>	Q2140	Sumatera	Acehnese	ace	32	0.68	31	0	ID	ID-SU	South
<input type="checkbox"/>	Q1823	Aceh	Acehnese	ace	4059	86.56	3246	1293	ID	ID-AC	South
<input type="checkbox"/>	Not Assigr		Acehnese	ace	598	12.75	449	158			
<input type="checkbox"/>	Q5328	Karachay-i	Adyghe	ady	2	6.25	1	0	RU	RU-KC	Easter
<input type="checkbox"/>	Q3734	Republic c	Adyghe	ady	14	43.75	11	0	RU	RU-AD	Easter
<input type="checkbox"/>	Q3680	Krasnodar	Adyghe	ady	2	6.25	2	0	RU	RU-KDA	Easter
<input type="checkbox"/>	Not Assigr		Adyghe	ady	14	43.75	4	7			
<input type="checkbox"/>	Q963	Botswana	Afrikaans	af	81	0.65	56	20	BW		Sub-S
<input type="checkbox"/>	Q258	South Afric	Afrikaans	af	11969	95.68	4290	399	ZA		Sub-S

INTERSECCIONS I INCREMENTS

Un cop obtingut el conjunt de dades de **CCC**, podem calcular les diferents interseccions amb altres grups per tal d'entendre la seva extensió i abast.

Entitats:

- Gent (Men and Women)
- Articles Geoubicats (Continents, world regions and countries)
- Articles sense interwiki links
- Tota la viquipèdia
- ...





What articles do languages have in common?

Culture Gap: la majoria d'articles de CCC no estan disponibles en tots els idiomes

Al voltant del 60% de forats de contingut entre llengües es deuen a CCC.



Com es cobreixen les llengües els articles de CCC?

Aquests panells són per obtenir una visió general sobre la cobertura (coverage) i la difusió (spread) de tot el CCC.

https://wcdo.wmflabs.org/ccc_spread

https://wcdo.wmflabs.org/ccc_coverage

De nou, sabem:

No podem omplir els forats... però podem centrar-nos en les llistes d'articles de Top CCC.

- Podeu aprofundir en el valor del vostre context.
- Espectacles de televisió, monuments, figures polítiques ...?...



Les grans Viquipèdies haurien de tenir l'objectiu de **cobrir el mínim de les cultures d'uns altres**. Estic més preocupat per la bretxa d'articles de Top CCC que tota la bretxa.

Les petites Viquipèdies haurien d'**intentar crear articles que puguin omplir les llistes d'articles Top CCC**. Aquest és el primer grup d'articles que el món hauria d'importar.

Tres panells de control

Countries Top CCC articles coverage by Catalan Wikipedia

This page shows some statistics that explain how well the first each Catalan Wikipedia Top 100 CCC articles (only the first 100) are spread across the other language editions.

Some languages are mapped to order to create lists for countries selection process, and later, they speaking territories), whose territory is covered by the language.

These lists are created by ranking plain CCC or geolocated articles from CCC, list of CCC articles with featured article distinction from CCC, list of CCC articles with categorized in Wikidata as men with last year and with most edits **Last Y.**

The following table is useful to assess sorted in alphabetic order by their Wikicode. The last column **Sum Covered Articles** present the percentage of articles covered by the language.

The challenge is to reach 100 articles covered by the language.

Country

Afghanistan (Persian CCC)	
Afghanistan (Lahnda languages CCC)	
Afghanistan (Pashto CCC)	
Afghanistan (Uzbek CCC)	

Languages Top 100 CCC articles lists coverage by Catalan Wikipedia

This page shows some statistics that explain how well the first each Catalan Wikipedia Top 100 CCC articles (only the first 100) are spread across the other language editions.

These lists are created by ranking the articles according to specific features and sometimes giving them weights. These different features are usually based on the content type (e.g. plain CCC or geolocated articles) or article characteristics (number of Bytes). The Top CCC articles lists are: list of CCC articles with most number of editors (**Editors**), list of CCC articles with featured article distinction (**Featured**), most bytes and references (weights: 0.8, 0.1 and 0.1 respectively), list of CCC articles with geolocation with most links coming from CCC, list of CCC articles with keywords on title with most bytes (**Bytes**), list of CCC articles categorized in Wikidata as women with most edits (**Women**), list of CCC articles categorized in Wikidata as men with most edits (**Men**), list of CCC articles created during the first three years and with most edits (**First 3Y.**), list of CCC articles created during the last year and with most edits (**Last Y.**), list of CCC articles with most pageviews during the last month (**Pageviews**), list of CCC articles with most edits in talk pages (**Discussions**).

The following table is useful in order to assess how known the Top 100 CCC articles from Catalan Wikipedia language are in the entire Wikipedia project. Languages are sorted in alphabetic order by their Wikicode, and columns present the number of articles from each list covered by the language. The last column **Sum Spread Articles** the overall sum of articles from the Top 100 of each list spread across a specific language.

The challenge is to reach 100 articles spread (Sum Spread Articles) across each other Wikipedia language edition!

Language	Wiki	Editors	Featured	Geolocated	Keywords	Women	Men	First 3Y	Last Y	Page views	Talk Edits	Sum Spread Articles	World Subregion
Abkhaz	ab	7%											
Acehnese	ace	13%											
Adyghe	ady	3/12											
Afrikaans	af	79%											
Akan language	ak	11/5											
Alemannic	als	83%											
Amparic	am	22%											
Arabic	ar	96%											
Old English	ang	78%											

Catalan Wikipedia Top 100 CCC article lists spread across the rest of Wikipedias

This page shows some statistics that explain how well the first each Catalan Wikipedia Top 100 CCC articles (only the first 100) are spread across the other language editions.

These lists are created by ranking the articles according to specific features and sometimes giving them weights. These different features are usually based on the content type (e.g. plain CCC or geolocated articles) or article characteristics (number of Bytes). The Top CCC articles lists are: list of CCC articles with most number of editors (**Editors**), list of CCC articles with featured article distinction (**Featured**), most bytes and references (weights: 0.8, 0.1 and 0.1 respectively), list of CCC articles with geolocation with most links coming from CCC, list of CCC articles with keywords on title with most bytes (**Bytes**), list of CCC articles categorized in Wikidata as women with most edits (**Women**), list of CCC articles categorized in Wikidata as men with most edits (**Men**), list of CCC articles created during the first three years and with most edits (**First 3Y.**), list of CCC articles created during the last year and with most edits (**Last Y.**), list of CCC articles with most pageviews during the last month (**Pageviews**), list of CCC articles with most edits in talk pages (**Discussions**).

The following table is useful in order to assess how known the Top 100 CCC articles from Catalan Wikipedia language are in the entire Wikipedia project. Languages are sorted in alphabetic order by their Wikicode, and columns present the number of articles from each list covered by the language. The last column **Sum Spread Articles** the overall sum of articles from the Top 100 of each list spread across a specific language.

The challenge is to reach 100 articles spread (Sum Spread Articles) across each other Wikipedia language edition!

Language	Wiki	Editors	Featured	Geolocated	Keywords	Women	Men	First 3Y	Last Y	Page views	Talk Edits	Sum Spread Articles	World Subregion
Afar	aa	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0	Sub-Saharan Africa
Abkhaz	ab	4%	1%	1%	1%	0%	1%	4%	0%	0%	4%	4	Western Asia
Acehnese	ace	3%	2%	1%	1%	0%	0%	2%	0%	0%	2%	4	South-eastern Asia
Adyghe	ady	2%	1%	1%	1%	0%	0%	2%	0%	0%	2%	2	Eastern Europe
Afrikaans	af	22%	7%	9%	6%	2%	12%	17%	0%	5%	15%	38	Sub-Saharan Africa
Akan language	ak	1%	0%	0%	0%	0%	0%	1%	0%	0%	1%	1	Sub-Saharan Africa
Alemannic	als	16%	4%	6%	4%	0%	11%	13%	0%	4%	11%	24	Western Europe
Amparic	am	18%	4%	10%	6%	0%	3%	18%	0%	2%	11%	23	Sub-Saharan Africa
Arabic	ar	71%	18%	79%	12%	10%	36%	79%	0%	13%	44%	196	Southern Europe
Old English	ang	6%	3%	2%	2%	0%	3%	7%	0%	0%	5%	9	Northern Europe

- **Languages Top CCC articles coverage**
- **Countries Top CCC articles coverage**



**WIKIPEDIA
ASIAN MONTH**

- **Languages Top CCC articles spread**

Catalan culture challenge

Com de bé es cobreixen els articles del Top CCC?

Aquests panells són útils per obtenir una vista general de la cobertura (coverage) i exportació (spread) dels Top CCC de i a altres llengües.

- **Languages Top CCC articles coverage**

https://wcdо.wmflabs.org/languages_top_ccc_articles_coverage/?lang=ca

- **Countries Top CCC articles coverage**

https://wcdо.wmflabs.org/countries_top_ccc_articles_coverage/?lang=ca

- **Languages Top CCC articles spread**

https://wcdо.wmflabs.org/languages_top_ccc_articles_spread/?lang=ca

Lang = wikicode

Passos futurs

Hi ha diverses possibilitats per millorar WCDO principalment des de l'anàlisi de dades i el punt de vista de la visualització. El projecte acaba d'establir la seva base i ara seria el moment de pensar en altres implementacions:

- Pageviews (quina és la magnitud de les pàgines vistes en CCC en tots els idiomes?).
- Cobertura temàtica (quins són els articles de Glam, monuments, natura, etc.? Podríem fer més llistes?).
- Monitoreig temporal de la bretxa/esforçoso (estem millorant aquest any que l'anterior? Estem dedicant esforços als llocs que ho requereixen?).
- Mapes i informes d'articles geolocalitzats (on es troben els buits?).
- Editors (qui són els editors multilingüe amb qui podria col·laborar?).
- Newsletter mensual amb part d'aquests resultats.

Crec que no hi ha res més Viquipedista que la multiculturalitat.

Una vegada reconeixem la mancança, col·laborem entre llengües i intercanviem el contingut amb d'altres.

Moltes gràcies!!!

Wikipedia Cultural Diversity Observatory (WCDO)

[<https://meta.wikimedia.org/wiki/WCDO>]

Marc Miquel {marcmiquel@gmail.com}

Usuari:marcmiquel

Viquitrobada, Manresa, 2018



Referències (si voleu aprofundir-hi)

Miquel-Ribé, M., & Laniado, D. (2016, July). Cultural identities in wikipeidias. In *Proceedings of the 7th 2016 International Conference on Social Media & Society* (p. 24). ACM.

Miquel-Ribé. M. (2017). *Identity-based motivation in digital engagement: the influence of community and cultural identity on participation in wikipedia* (Doctoral dissertation, Universitat Pompeu Fabra).

Miquel-Ribé, M., & Laniado, D. (2018). Wikipedia Culture Gap: Quantifying Content Imbalances Across 40 Language Editions. *Frontiers in Physics*, 5, 12. (CC BY) Open Access.

Agraïments a:

