

Minoritized Languages and Missing Languages:

An Opportunity to Increase Cultural Diversity in Wikipedia

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

Marc Miquel, PhD

{marcmiquel@gmail.com}

Username:marcmiquel

Pompeu Fabra University, Barcelona, Catalonia

Amical Wikimedia (Catalan Wikipedia)

WG Diversity – Strategy 2030



WIKIMANIA
STOCKHOLM

AUGUST 14-18, 2019



The Problem

Wikipedia project does not reflect enough the world's cultural diversity.



What can we do?

Wikipedia Cultural Diversity Observatory (WCDO).

“a joint space for **researchers, developers and activists** to study and **fight against the knowledge gaps** and increase cultural diversity in contents”.



<http://wcd0.wmflabs.org>

Problem of lack of Cultural Diversity in Wikipedia

Why is Wikipedia failing at gathering the human cultural diversity?

1) Concepts from one context not represented in their native language because they do not have a Wikipedia.

We should look for new potential Wikipedias.

2) Concepts from one context not represented in their native language even though it has a Wikipedia language edition.

We should encourage these Wikipedias to represent their context.

3) Concepts are represented but not shared. We won't talk about this.

We should encourage and facilitate content translation.

Why is Wikipedia failing at gathering the human cultural diversity?

1) Concepts from one context not represented in their native language because they do not have a Wikipedia.

We should look for new potential Wikipedias.

(1) Look for Potential Wikipedias

There exist approximately 7000 languages according to Ethnologue (SIL).

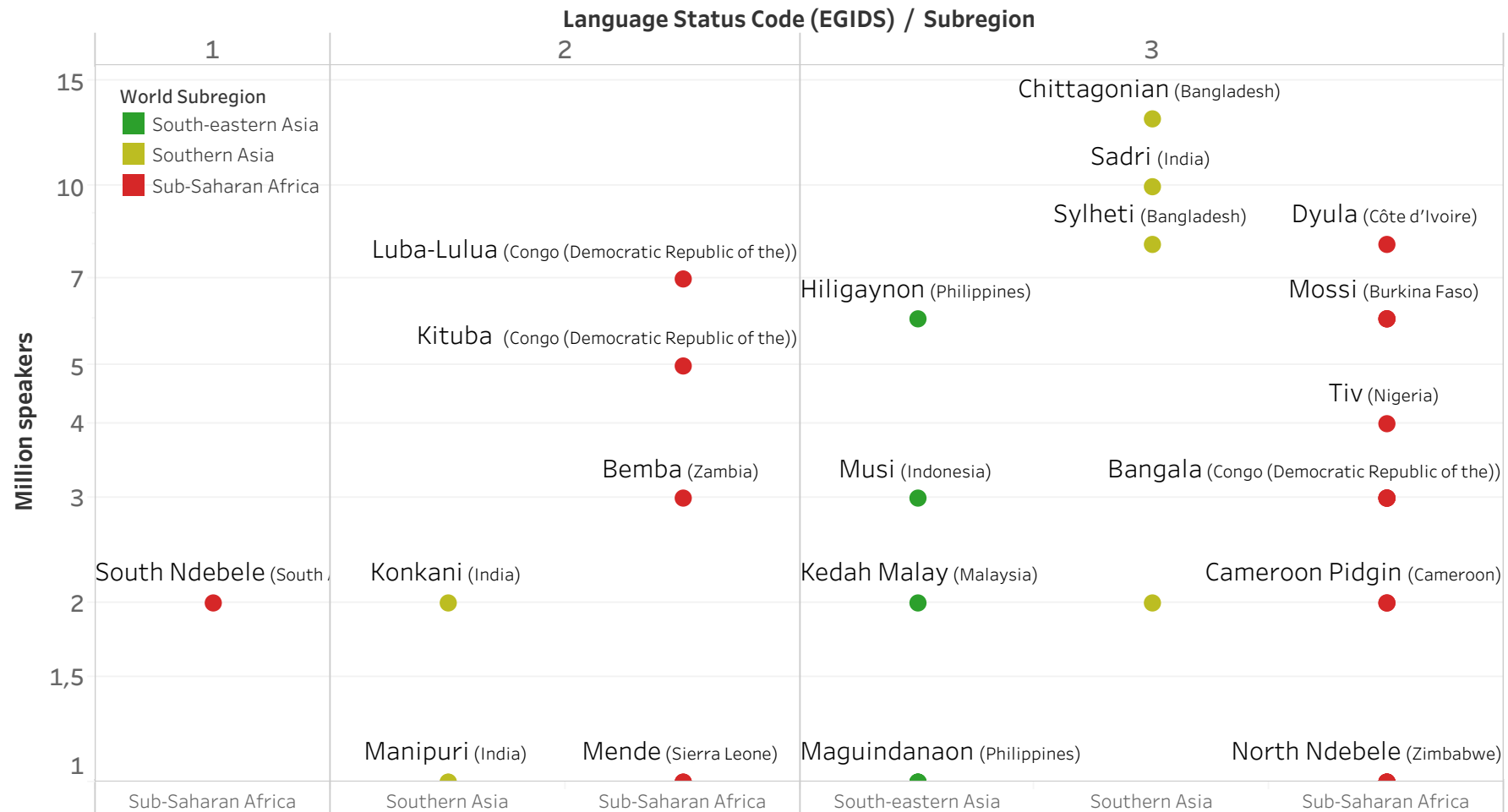
We need to detect potential languages that can easily become Wikipedias.

But which?

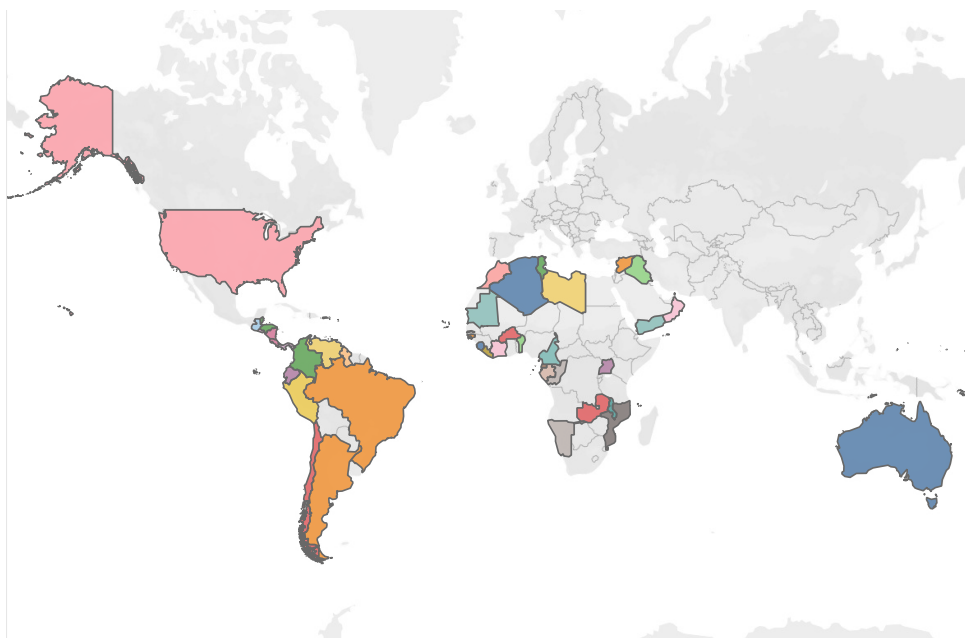
EGIDS (language social status) sets some priority along the number of speakers.

Potential new Wikipedias

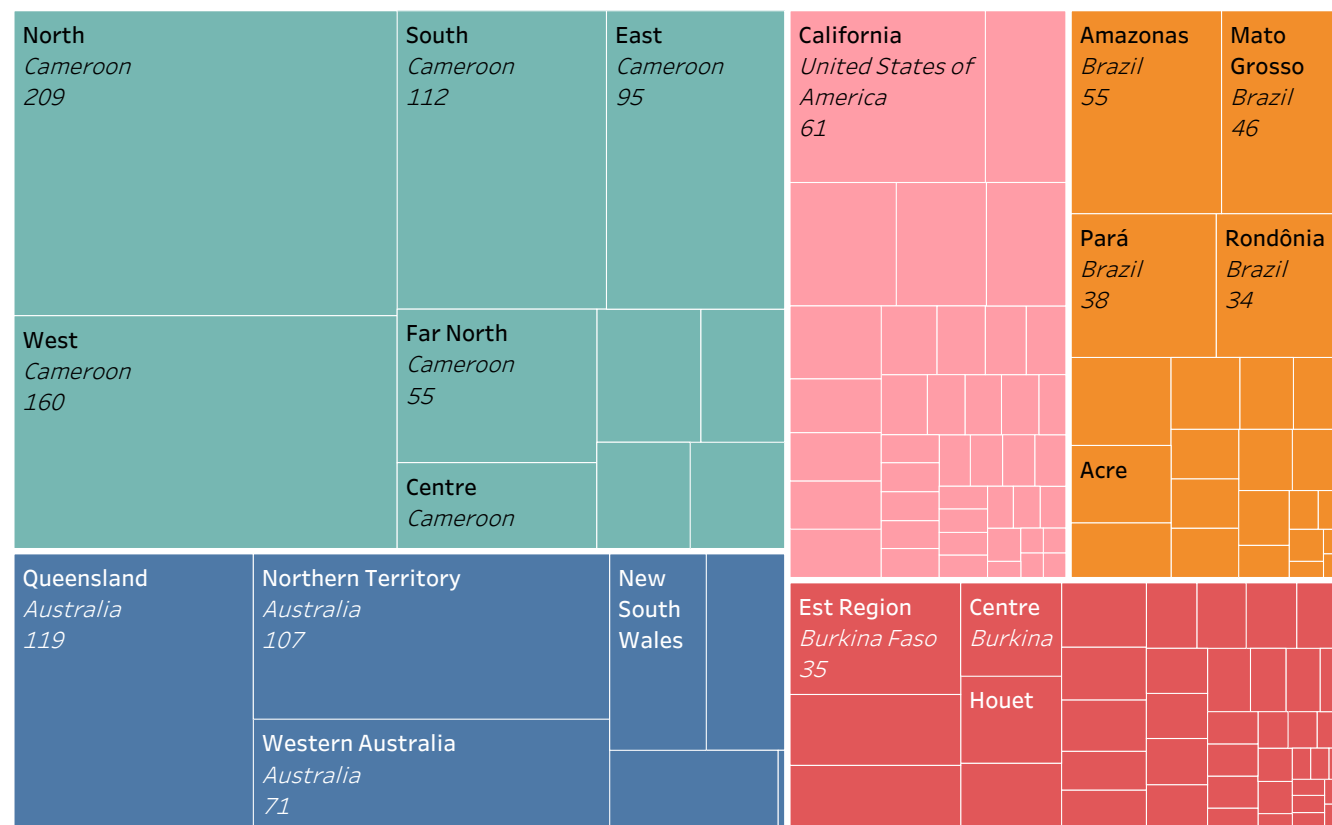
We need to detect potential languages that can easily become Wikipedias with million speakers and the language status code (EGIDS).



- **Missing territories (not covered by a Wikipedia of an indigenous language)**



Countries not covered by an indigenous language with a Wikipedia.



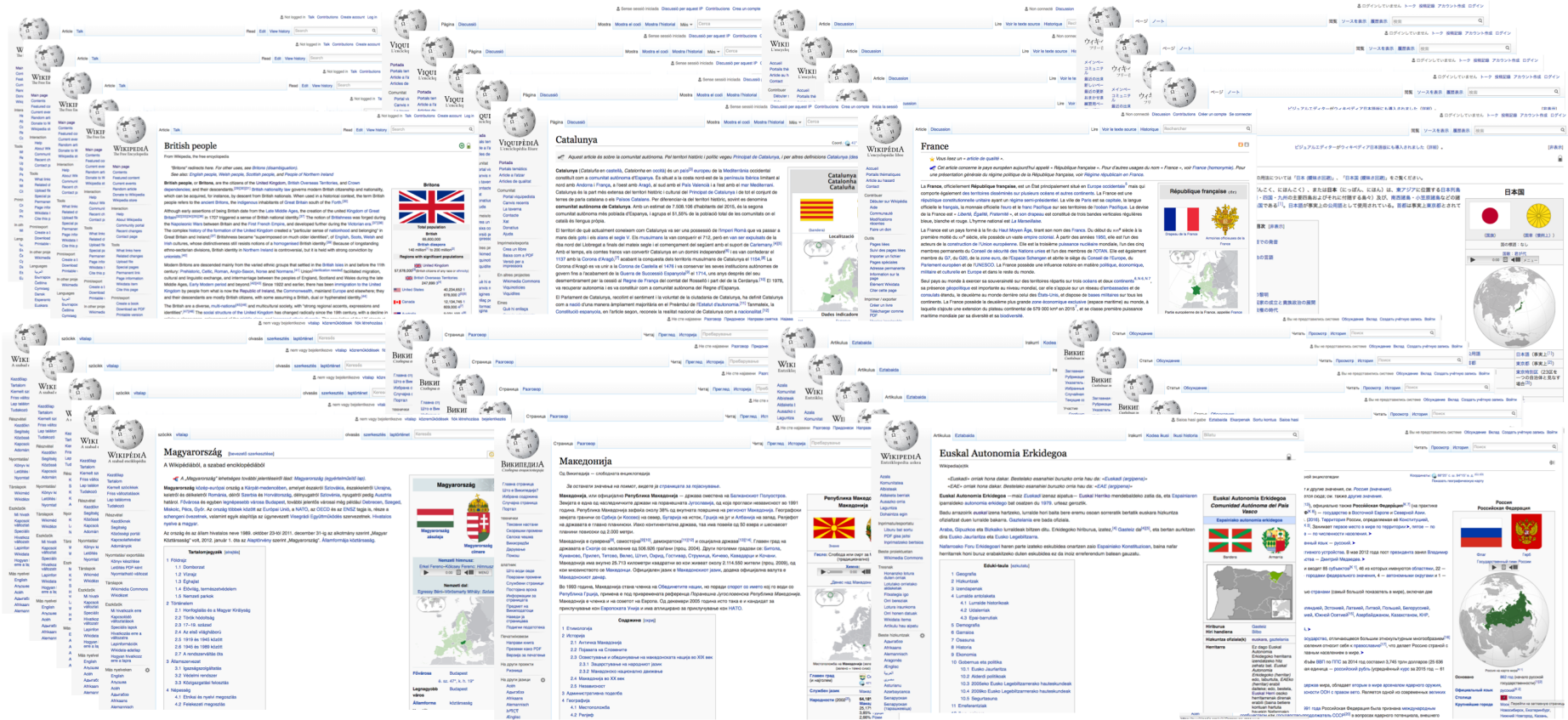
Top five countries (colour) and subdivisions not covered by a Wikipedia of an indigenous language by the number of indigenous languages (size).

Why is Wikipedia failing at gathering the human cultural diversity?

2) Concepts from one context not represented in their native language even though it has a Wikipedia language edition.

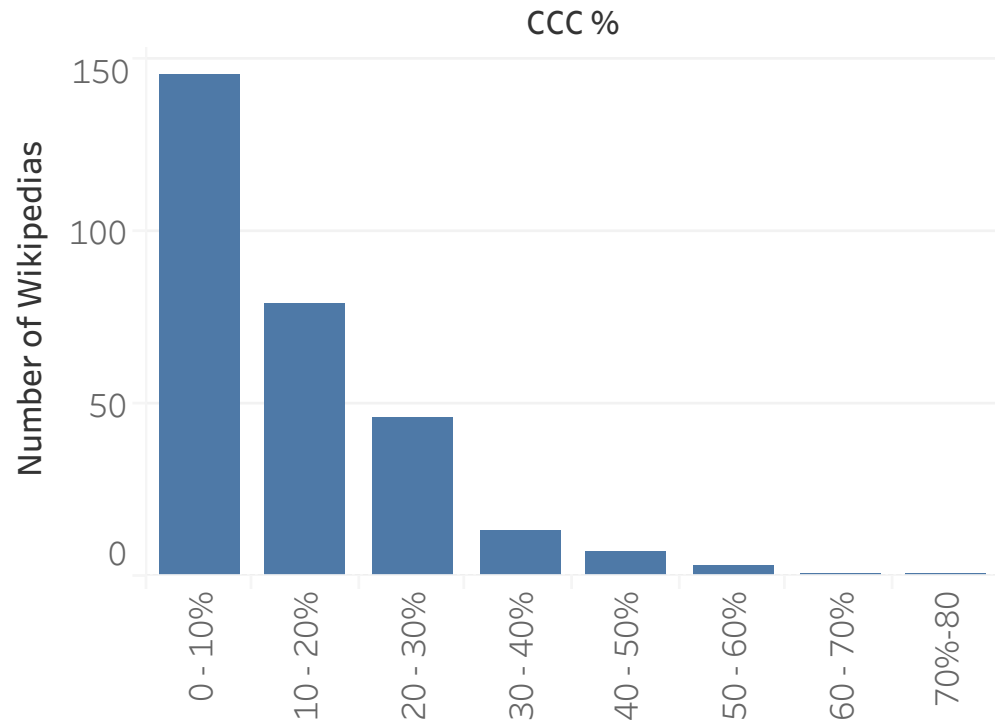
We should encourage these Wikipedias to represent their context.

**(2) Encourage cultural context
representation in small languages**



For each Wikipedia language edition, we aim at selecting the **Cultural Context Content (CCC)**, i.e. traditions, language, politics, agriculture, biographies, places, events, etcetera, related to the territories where the language is spoken.

We have many Wikipedias with a small CCC



- 92 with < 100 art. geolocated in their territories.
- 156 with < 100 art. including keywords such as the language name or territory name in their title.
- 197 with < 100 art./men biographies related to their territories.
- 197 with < 100 articles/women biographies related to their territories.

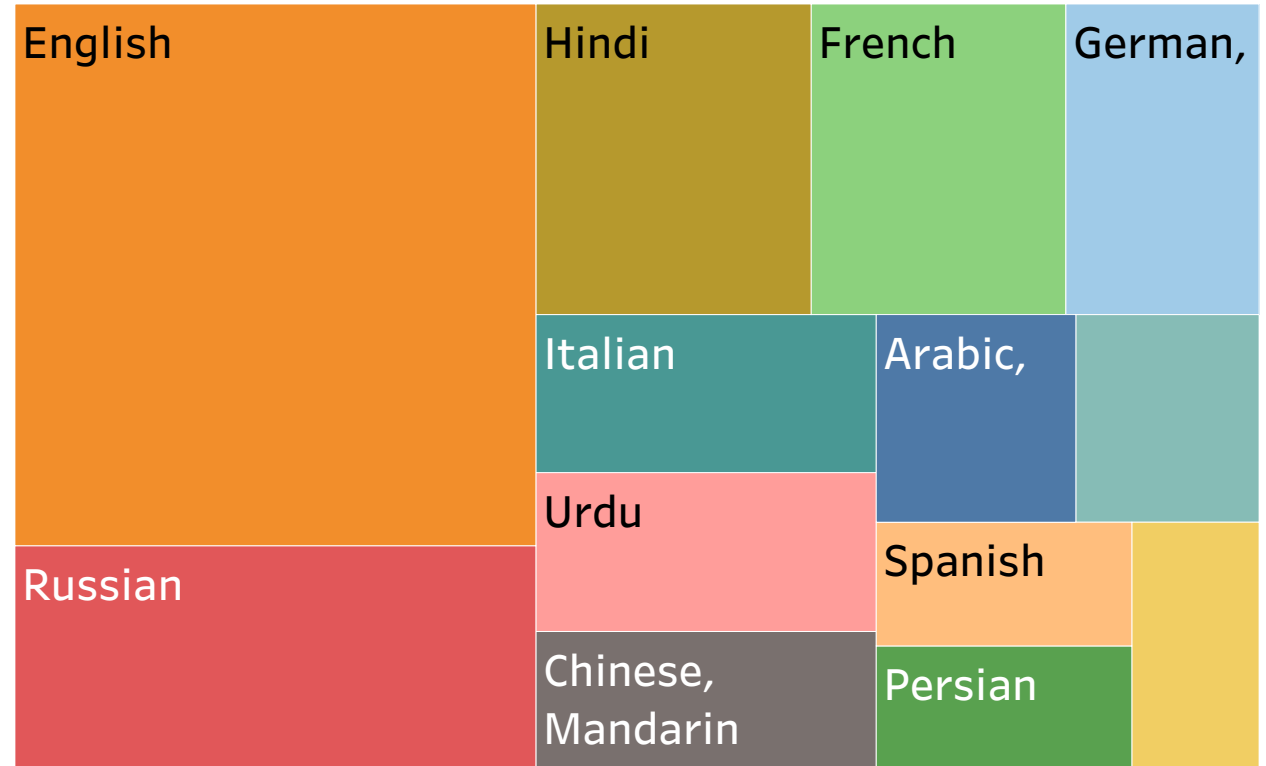
145 Wikipedias CCC is below 10% of their content.

[https://wcdo.wmflabs.org/list_of_wikipedias_by_cultural_context_content]

- **Minoritized and minority languages**

253 out of the 302 Wikipedia languages are from languages which coexist with other languages with a higher social status.

English and Russian are the languages which overlap most with other Wikipedia languages (87 and 37) because they are official at country level.



The inverse unusual situation

In some African languages, cultural context content tends to be represented in English or French rather than in their native Wikipedia.

For instance, **Luganda Wikipedia** (from Uganda) has a very low CCC (**just 3.15%**). There is an opportunity to give a digital life to the language through Wikipedia.



- We can create lists of **missing CCC articles** in minoritized languages that exist in higher status languages and encourage them to create them.

For example: **Luganda language** coexists with **English** as a higher status language (“Status 3 wider communication” and “Status 1 national” correspondingly).

Luganda Wikipedia (1171 articles) lacks articles about:

- Geography
- Traditions
- Politicians
- Etc.

Not logged in | Talk | Contributions | Create account | Log in

Article | Talk | Read | Edit | View history | Search Wikipedia

Yoweri Museveni

From Wikipedia, the free encyclopedia

The **neutrality of this article is disputed**. Relevant discussion may be found on the **talk page**. Please do not remove this message until conditions to do so are met. *(January 2014)* [\(Learn how and when to remove this template message\)](#)

Yoweri Kaguta Museveni (ⓘ pronunciation (help·info); born 15 September 1944) is a Ugandan politician who has been **President of Uganda** since 1986. Museveni was involved in rebellions that toppled notorious Ugandan leaders **Idi Amin** (1971–79) and **Milton Obote** (1980–85) before capturing power in the 80s. In the mid to late 1990s, Museveni was celebrated by the **West** as part of a **new generation of African leaders**. During Museveni's presidency, Uganda has experienced relative peace and significant success in battling **HIV/AIDS**. At the same time, Uganda remains a country suffering from high levels of corruption, unemployment and poverty. Museveni's presidency has been marred by involvement in the **civil war in the Democratic Republic of the Congo** and other **Great Lakes region** conflicts; the rebellion in Northern Uganda by the **Lord's Resistance Army** which caused a drastic

Yoweri Museveni

Museveni in July 2012

9th President of Uganda

President of Uganda does exist

Not logged in | Talk | Contributions | Create account | Log in

Article | Talk | Read | Edit | View history | Search Wikipedia

Ruhakana Rugunda

From Wikipedia, the free encyclopedia

Ruhakana Rugunda (born 7 November 1947) is a Ugandan politician who has been **Prime Minister of Uganda** since 2014. A physician by profession, he held a long series of Cabinet posts under President **Yoweri Museveni** beginning in 1986. He served as Uganda's Minister of Foreign Affairs from 1994 to 1996 and as Minister of Internal Affairs from 2003 to 2009. Subsequently, he was Permanent Representative to the **United Nations** from 2009 to 2011 and Minister of Health from 2013 to 2014.

He was appointed as Prime Minister on 18 September 2014. He replaced **Amama Mbabazi**, who was dropped from the Cabinet.^[1]

Contents [hide]

- 1 Early career
- 2 Political career
- 3 Personal
- 4 See also
- 5 References

Rugunda at the World Trade Organization Ministerial Conference of 2015

10th Prime Minister of Uganda

Incumbent

Assumed office
18 September 2014

President **Yoweri Museveni**

Prime Minister of Uganda does not exist

Missing CCC Articles

Select the Wikipedia

Target Language

Target Country (Optional)

Target Region (Optional)

Zulu (zu)



Select a country (optional)

Select a region (optional)

Filter by content

Search Type

CCC Segment

Topic

Missing articles



Select a CCC Segment

Select a topic

Choose source of content

Source language(s)

Order by feature

Limit the results

Source language



Number of Inlinks



100

QUERY RESULTS!

N°	Language	Title	Number of Editors	Number of Pageviews	Number of Inlinks	Number of Bytes	Label (Lang)	Qitem
1	it	Nazionale di calcio del Sudafrica	179	137	1175	63400	en South Africa national football team	Q170327
2	ar	كأس العالم 2010	385	498	1172	233696	en 2010 FIFA World Cup	Q176883
3	de	Fußball-Weltmeisterschaft 2010/Qualifikation	0	39	818	210877	en 2010 FIFA World Cup qualification	Q193596
4	af	Universiteit van Pretoria	32	10	533	12450	en University of Pretoria	Q604444
5	pl	Liesel Huber	125	1	501	80780		Q53565
6	nl	Wayne Ferreira	24	3	487	46170	en Wayne Ferreira	Q369228
7	af	Pan Africanist Congress of Azania	18	5	455	15702	en Pan Africanist Congress of Azania	Q775460
8	af	United Democratic Movement	14	6	452	7226		Q1788070
	af	Afrika Christen Demokratiese Party	25	24	451	12455		Q268613
undo	it	Brent Haygarth	14	0	444	8847		Q909744
11	it	Grant Stafford	14	0	442	9201	en Grant Stafford	Q1543225

* Alpha version

We should encourage them not only to build a Wikipedia,

**But to create their most unique and specific knowledge about their context:
few lists of 100 articles on different topics can really make a difference.**



Arctic char fishing. By Ansgar Walk.

We need to value their knowledge and difference.

Why is Wikipedia failing at gathering the human cultural diversity?

3) Concepts are represented in one language but not shared across the other.

This is going to be in another session.

Thank you very much!

Wikipedia Cultural Diversity Observatory (WCDO)

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

Marc Miquel

{marcmiquel@gmail.com}

Username:marcmiquel

Pompeu Fabra University, Barcelona, Catalonia

Amical Wikimedia (Catalan Wikipedia)

Wikimedia Foundation – Project Grant



References (if you want to know more)

Miquel-Ribé, M., & Laniado, D. (2019). Wikipedia Cultural Diversity Dataset: A Complete Cartography for 300 Language Editions. *Proceedings of the 13th International AAAI Conference on Web and Social Media. ICWSM. ACM.*

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.

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. (2016). Cultural identities in wikipeidias. In *Proceedings of the 7th 2016 International Conference on Social Media & Society* (p. 24). ACM.

