



**WIKIPEDIA**  
The Free Encyclopedia

# Multilingual User Generated Content at Wikipedia

---

**Alolita Sharma**

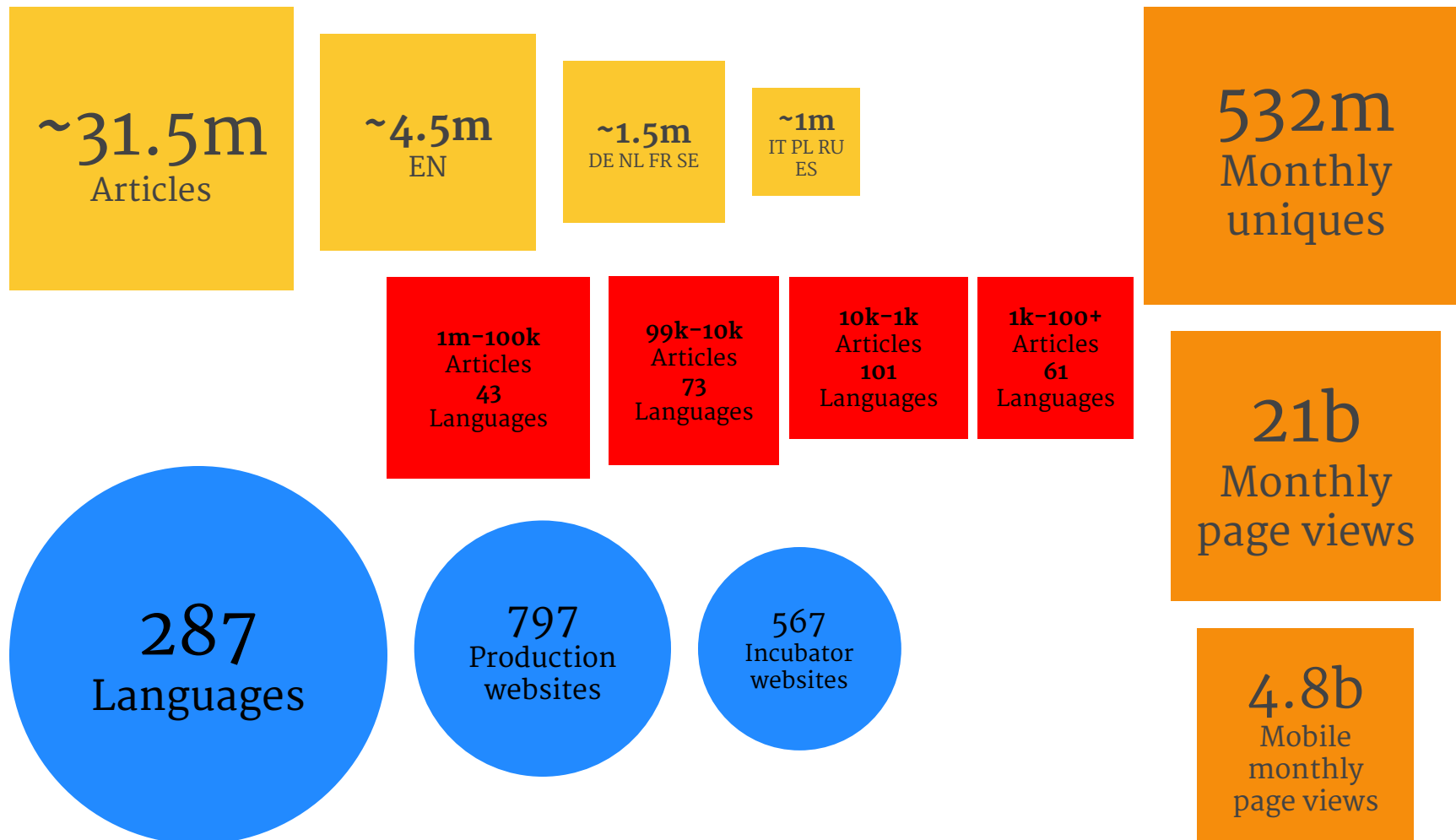
Director of Language Engineering  
Wikimedia Foundation  
[asharma@wikimedia.org](mailto:asharma@wikimedia.org)

**New Horizons for the Multilingual Web – W3C Multilingual Web Conference**

ETSIT-UPM, Madrid, May 7 2013 #mlwmadrid  
CC-BY-SA 3.0

---

# Wikipedia-Scale



# Wikipedia: Growth by Region



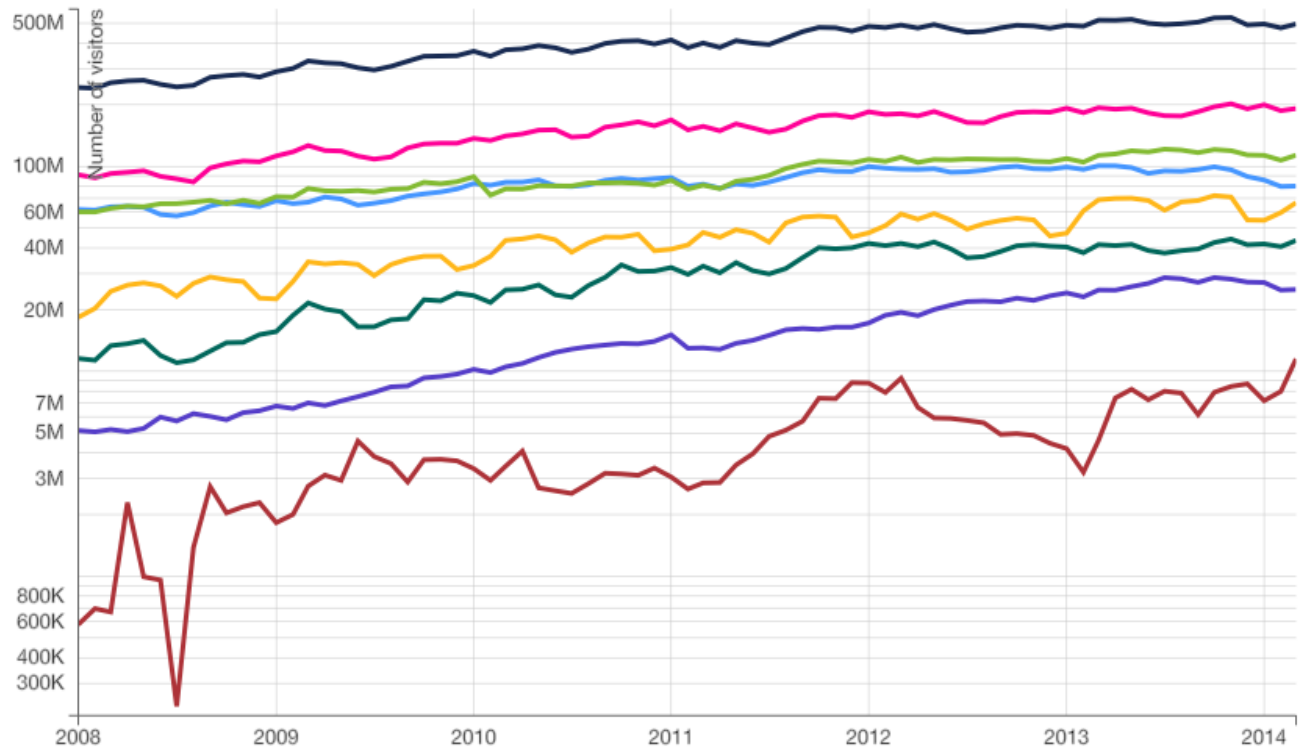
495.18 Million

Mar 13 – Mar 14 -4.33%  
Feb 14 – Mar 14 4.44%

Unique Visitors per Region (comScore)

Nov 2013

World	532.70M
China	8.44M
Europe	202.10M
India	28.16M
Latin America	70.86M
Middle East/Africa	44.19M
North America	96.22M
Asia Pacific	119.32M



# Wikipedia: Mobile Growth



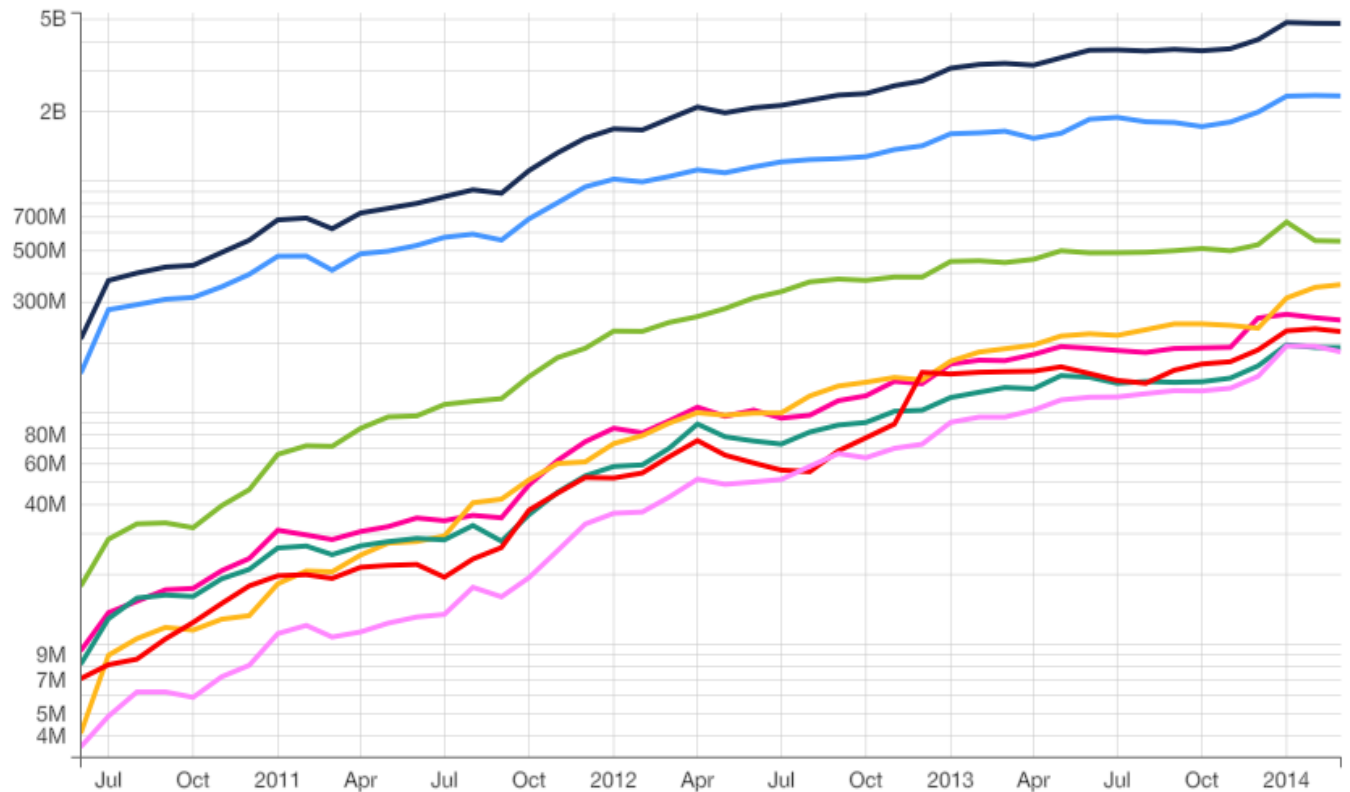
4.80 Billion

Page Views to Mobile Site

Mar 13 – Mar 14 48.78%  
Feb 14 – Mar 14 -0.19%

Mar 2014

<b>Total</b>	4.80B
<b>English Mobile</b>	2.33B
<b>Japanese Mobile</b>	550.63M
<b>German Mobile</b>	250.99M
<b>Spanish Mobile</b>	357.30M
<b>French Mobile</b>	190.97M
<b>Russian Mobile</b>	223.95M
<b>Italian Mobile</b>	183.09M



# Wikipedia today



1 000 000+ articles

No	Language	Language (local)	Wiki	Articles	Total	Edits
1	English	English	en	4,507,774	32,821,477	716,354,339
2	Dutch	Nederlands	nl	1,774,505	3,205,497	42,212,245
3	German	Deutsch	de	1,715,539	4,746,824	135,977,632
4	Swedish	Svenska	sv	1,618,424	3,610,936	26,566,975
5	French	Français	fr	1,502,818	6,470,166	105,861,687
6	Italian	Italiano	it	1,119,723	3,655,232	71,063,857
7	Russian	Русский	ru	1,110,690	3,739,694	75,205,855
8	Spanish	Español	es	1,098,911	4,559,462	79,701,638
9	Polish	Polski	pl	1,043,983	2,062,360	39,395,536

# Wikipedia today



## 100 000+ articles

No ↕	Language ↕	Language (local) ↕	Wiki ↕	Articles ↕	Total ↕	Edits ↕
10	Waray-Waray	Winaray <a href="#">↗</a>	war	959,771 <a href="#">↗</a>	1,980,590	4,867,190
11	Japanese	日本語 <a href="#">↗</a>	ja	907,711 <a href="#">↗</a>	2,539,048	52,609,371
12	Cebuano	Sinugboanong Binisaya <a href="#">↗</a>	ceb	892,119 <a href="#">↗</a>	1,878,479	4,537,168
13	Vietnamese	Tiếng Việt <a href="#">↗</a>	vi	888,199 <a href="#">↗</a>	2,298,929	16,384,752
14	Portuguese	Português <a href="#">↗</a>	pt	826,676 <a href="#">↗</a>	3,509,946	39,885,533
15	Chinese	中文 <a href="#">↗</a>	zh	766,616 <a href="#">↗</a>	3,394,404	32,598,519
16	Ukrainian	Українська <a href="#">↗</a>	uk	498,099 <a href="#">↗</a>	1,468,478	14,433,374
17	Catalan	Català <a href="#">↗</a>	ca	427,345 <a href="#">↗</a>	1,066,616	13,707,545
18	Norwegian (Bokmål)	Norsk (Bokmål) <a href="#">↗</a>	no	418,789 <a href="#">↗</a>	989,380	14,188,631
19	Persian	فارسی <a href="#">↗</a>	fa	388,658 <a href="#">↗</a>	2,222,566	17,305,842
20	Indonesian	Bahasa Indonesia <a href="#">↗</a>	id	351,399 <a href="#">↗</a>	1,366,544	8,890,468
21	Finnish	Suomi <a href="#">↗</a>	fi	346,202 <a href="#">↗</a>	927,727	14,884,230
22	Czech	Čeština <a href="#">↗</a>	cs	294,343 <a href="#">↗</a>	759,580	11,825,690
23	Arabic	العربية <a href="#">↗</a>	ar	278,250 <a href="#">↗</a>	1,661,756	15,731,477
24	Korean	한국어 <a href="#">↗</a>	ko	275,927 <a href="#">↗</a>	899,566	14,213,400
25	Malay	Bahasa Melayu <a href="#">↗</a>	ms	260,364 <a href="#">↗</a>	682,778	3,780,100

# Who are our users

---



User generated content (UGC) is related to level of online activity of language communities.

**Early Adopters:** Large languages with large active online communities generating lots of content e.g. Latin languages (English, German, Dutch, French)

**Next Generation:** Large languages with small online communities are generating very little content e.g. Indic languages, Right-to-left languages, CJK languages

**Long tail** languages with tiny but passionate communities starting with little content e.g Native American Indigenous languages, Newari

---

# Growing Content Contributions

---



Content has to grow at the same pace as rich delivery platforms

Access to content has to be free and pervasive

Virtuous cycle of contributions

Rich language tools and language assets for end users

The tablet is the platform

Let a thousand language web applications grow for  
contributing and consuming content  
to Wikipedia and other websites

---



# Challenges

---



Accessible and open content

High quality content

Broken user experience for multilingual users

Inadequate language tools for Web and Mobile

Lack of reference data corpora

Growing contributor communities

---



# What is Wikipedia doing

---

## **Language Selection**

Universal Language Selector to set language preferences for display UI, fonts, input tools  
Smart language selector search handling multiple scripts

## **Web Fonts**

High-quality web fonts for 63 languages with 81 variants

## **Input Tools**

Easy-to-use 139 input methods for 64 languages  
Onscreen keymaps

## **Internationalization**

Javascript and PHP i18n support for grammar, plurals, gender

## **Content Translation**

Content translation platform integrated in Wikipedia, side-by-side translation editor  
Machine translation, Translation Memories, Dictionaries, Glossaries, Wikidata

## **Software UI and Message Localization**

Translation platform, side-by-side proofreading editor with translation aids  
Crowdsourced web platform - [translatewiki.net](https://translatewiki.net)

## **Wikipedia Zero - Access for All**

---

# Where we are headed

---

As Wikipedia turns 14, it has become the most significant open content platform of this century.

Content Commons for the Web

Generate rich high-quality user content

Deliver first-class multilingual user experience

Engage new generation of users

Be mobile, be everywhere

Commoditize language software

Keep the Web Open and Free

---

# Collaborate to make the Web Multilingual

---



Empower web and mobile platforms with language tools  
Create and release high quality language fonts and typing tools  
Collaborate with us to develop open tools, platforms and communities

Enable free access to content

Grow content communities

Build mobile web application developer ecosystem  
Seed language applications to enable content contributions

Report problems you encounter when you're reading or editing  
Wikipedia in your language

---

# Thank you!

---

May 7: 10:45–11:15

Best Practices on the Design of Translation

Pau Giner, David Chan and Santhosh Thottingal

May 8: 14:15–15:00

Panel 1: Using Wikipedia for multilingual web content analytics across 287 languages.

May 8: 15:30–16:15

Panel 2: Growing Wikipedia editing with intelligent multi-language suggestion lists for article translation as well as other techniques and tools.

---