Wikipedia Administrative Pages Analytics

Marc Miquel i Ribé, PhD
User:marcmiquel / Catalan Wikipedia
marcmiquel@gmail.com

WIKIMEDIA MEETING 2022

Hi everyone!

Welcome to this presentation. It is called "Wikipedia Administrative Pages Analytics"

**Hello! My name is Marc Miquel,
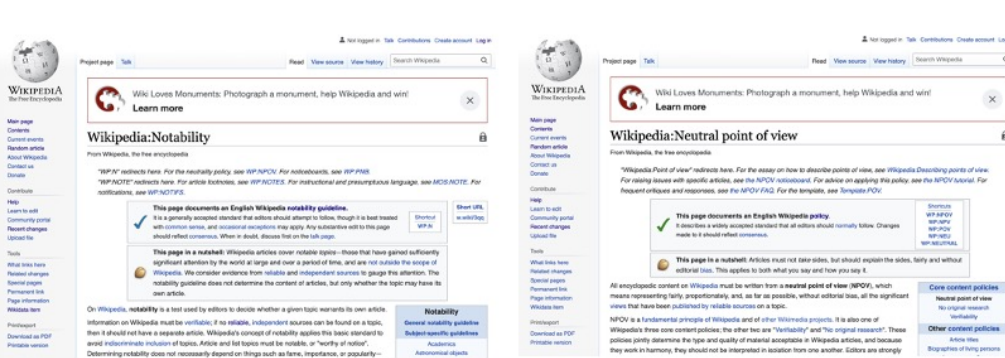I'm a Wikimedia researcher.**

Wikipedia Administrative Pages.

My name is Marc Miquel and I'm a Wikimedia researcher. I joined the Movement in 2011.

You may possibly know me from past projects. I have been doing several things on editor engagement, diversity and community health.

Today I want to talk about Administrative Pages.

1. Introduction

These are admin pages and they were created <u>twenty years ago.</u>

They are the backbone: what has enabled us to make Wikipedia successful.

So, we owe them a lot. We were an experiment that worked.

However, what is old, sometimes it is not maintained, it is stiff,...closed to new ideas and to new people.
That's not Wikipedia, necessarily, but in general.

2. Motivation

Welcome to the Jungle. **We need to research this.**

For Newcomers. Then, for all Wikipedians.

https://en.wikipedia.org/wiki/Wikipedia:Administration
https://en.wikipedia.org/wiki/Category:Wikipedia_administration

Image author: James Martins

Admin pages are the jungle. For two reasons:

They are hard to get through as an editor (newcomers); they have not been explored by researchers.

If admin pages are the jungle, we need to make them more welcoming and inclusive to all editors.

I do not want anybody to misinterpret me. I mean, they grow organically like the rest of the Wikipedia. When there is a need, when somebody remembers. But, the organic growing may create unnecessary clutter, some things not up to date, and not written from the perspective of the reader.

As I said, they are the jungle also because they are not explored from the research perspective. We have an opportunity.

Little Research on Admin Pages

1. Numerous and abundant rules and policies have been created that serve a wide variety of functions to maintain Wikipedia (Butler and Pike, 2008).

2. Newcomers find it difficult to have their contributions to policies accepted with an impact on their retention (Halfaker et al. 2013).

3. Participants (newcomers) to Wikipedia edit-a-thons experience moments of frustration when navigating a complex bureaucracy of policies and procedures (Gluza et al. 2021).

4. Similar rule-making activity across the five communities replicates and extends prior work on English language Wikipedia alone (Hwang and Shaw, 2022).

https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics/
Research_on_Wikipedia_Administrative_Pages

There is very little research on admin pages.
This shocked me, since Wikipedia is one of the most research objects on the Internet, if not the most.

Very few academic and scientific studies pay attention to admin pages: mostly on policies.

We know that, in 2008, Wikipedia was seen as a developed bureaucracy.
That later on, in 2013, when the decline in number of editors started, Halfaker saw that newcomers' contributions to policies were not accepted as much as those made by more seasoned editors.
Newcomers often experience moments of frustration navigating policies and procedures.
And the development of policies tends to be similar across languages and extends and replicates what has been done in English Wikipedia.

Little Research on Admin Pages

Wikipedia is one of the most researched objects. With studies about many aspects, which could also be explored for admin pages:

- **Topical analysis** (Halavais and Lackaff, 2008; Kittur et al. 2009).

Content gaps (Miquel-Ribé and Laniado, 2021; Redi et al., 2020).

- **Content biases and diversity** (Callahan and Herring, 2011; Beytía and Wagner, 2022)
- **Contextualization and cross-language analysis** (Hecht and Gergle, 2010; Hecht, 2013, Samoilenko et al. 2016).
- **Content quality models** (Lewoniewski et al. 2019; Halfaker and Geiger, 2020).
- **Content readability and style** (Setia et al. 2021, Liu et al. 2021).
- **Participation, and inequality** (Ortega and Robles, 2008; Shaw and Hargittai, 2018).
- **Cooperation and peer-to-peer communication** (Wilkinson and Huberman, 2007; Kittur and Krau, 2008).
- **Readership** (Warncke-Wang et al. 2015; Lehmann et al. 2014).
- **Vandalism** (Shachaf and Hara, 2010; Adler et al. 2011).

To have an idea, research on Wikipedia encompasses a wide variety of topics or areas.

Some of these could be explored for admin pages. Today they are not.
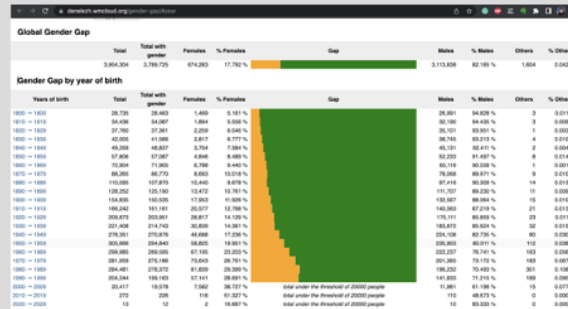
Gender Gap Dashboard (Denelezh, 2022)
https://denelezh.wmcloud.org/gender-gap/#year

For this reason, I believe we can say there is a "knowledge gap". A topic we do not know about.

Admin pages are a type of content. They are the content that regulates Wikipedia.

Maybe because we are "proud" of the result "the product", we hide them, because they are "the process". People do not study them.

Because of this, there is no framework that organizes the knowledge about them or analyzes their characteristics.

Differently than with content gaps, we know very little about what is available in every language, and we do not keep track of how they grow.

All in all, we lack a systematic understanding of administrative pages.

Wikipedia Administrative Pages Analytics (WAPA) is a research project to measure and improve maintenance and inclusion in Wikipedia administrative pages across Wikipedia language editions.

"We envision a community that regularly consults analytical tools to make informed decisions on which pages to ***create, improve or delete*** *and, also, to modify the course of current processes in order to be* ***inclusive*** *to all Wikimedians."*

Project Principles: efficiency, maintenance and inclusion.

https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics

To fill this gap, I want to present the initiative Wikipedia Administrative Pages Analytics (WAPA). It is a research project that sets the first stone to measure and improve maintenance and inclusion in admin pages across language editions.

We would like to envision a community that is aware of the state of the admin pages, maintains them and engages all types of Wikipedians in updating and growing them.

For this reason, the project principles are efficiency, maintenance, and inclusion.

2. Project Philosophy

"Let chaos (disorganized admin pages) reign,
and in chaos (with analytics) I will reign."*

Andrew Grove
* parentheses mine

https://en.wikiquote.org/wiki/Andrew_Grove

There is a quote by Andrew Grove, an Hungarian-American businessman
and engineer who served as the third CEO of Intel, that I like.

It says: "let chaos reign, and in chaos, I will reign".
I think it summarizes well the idea of the project.

We do not need to cut down the jungle, we just need the analytical tools,
the aerial view, to make sense of it.

3. Objectives

Objectives

1. Understand the state of development of the admin pages.
2. Classify the admin pages according to different types.
3. Create applications in order to support editing the admin pages.

In Admin Pages Analytics, I think we need to aim at three objectives.

1. Understand the state of development of the admin pages.

2. Classify the admin pages according to different types.

3. Create applications in order to support editing the admin pages.

This is what I will try to communicate in this presentation as well.

4. Approach

In order to reach the three objectives, we need to:

1. First, collect the admin pages and analyze them (5.1).
2. Then, classify them into different types of pages (5.2).
3. Finally, create web applications with tools (5.3).

We are using the dumps provided at dumps.wikimedia.org.
The data processing and hosting of the applications happens at Toolforge.

https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics/Methodology
https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics/Technical_Documentation
https://github.com/marcmiquel/WAPA

In order to reach the three objectives, we need to:

1. First, collect the admin pages and analyze them (5.1).

2. Then, classify them into different types of pages (5.2).

3. Finally, create web applications with tools (5.3).

We are using the dumps provided at dumps.wikimedia.org.

The data processing and hosting of the applications happens at Toolforge.

We are using Python, Pandas and Sqlite3.

5. Results
5.1 Collecting the admin pages

What is an admin page?

Administrative Namespaces (No content or discussion oriented)

"Wikipedia requires a certain amount of administration and governance in order to further the project's goals.

To achieve Wikipedia's purpose, a wide range of **administrative pages** are made available in various namespaces which enumerate the various protocols and conventions created and implemented by community consensus for English Wikipedia."

https://en.wikipedia.org/wiki/Wikipedia:Administration

Wikipedia data structure			
Namespaces			
Subject namespaces		**Talk namespaces**	
0	(Main/Article)	Talk	1
2	User	User talk	3
4	Wikipedia	Wikipedia talk	5
6	File	File talk	7
8	MediaWiki	MediaWiki talk	9
10	Template	Template talk	11
12	Help	Help talk	13
14	Category	Category talk	15
100	Portal	Portal talk	101
118	Draft	Draft talk	119
710	TimedText	TimedText talk	711
828	Module	Module talk	829

So, collecting the admin pages.
But what is an admin page?

It is basically the type of page that helps Wikipedia achieve its purpose. Protocols, conventions, help pages, etcetera.
They are not content, nor personal or discussion spaces. They help in creating the content.

Administrative Namespaces (No content or discussion oriented)
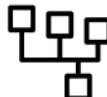
Template (ns10)   Categories (ns14)   Portals (ns100)

Wikipedia (ns4)   Help (ns12)

These two namespaces are the admin pages.

These are the namespaces that are not directly related to content or discussion.

Most typically, admin pages are namespace 4 (Wikipedia). We can also safely consider namespace 12 (Help) admin pages.

Admin Pages

Language	Admin P. (%)	Content (%)	Admin Pages	Content	Sum
English	17,19	82,81	1,360,647	6,556,492	7,917,139
German	3,76	96,24	106,624	2,729,665	2,836,289
Swedish	1,09	98,91	28,091	2,553,869	2,581,960
French	2,34	97,66	58,892	2,459,001	2,517,893
Russian	3,91	96,09	75,661	1,857,186	1,932,847
Spanish	2,04	97,96	37,615	1,806,193	1,843,808
Italian	8,07	91,93	155,628	1,773,667	1,929,295
Polish	5,78	94,22	94,264	1,537,322	1,631,586
Japanese	9,88	90,12	147,390	1,345,075	1,492,465
Arabic	3,06	96,94	37,530	1,188,062	1,225,592
Portuguese	9,78	90,22	118,805	1,095,400	1,214,205
Catalan	1,18	98,82	8,499	711,831	720,330
Romanian	4,62	95,38	20,997	433,727	454,724
Armenian	2,51	97,49	7,563	293,417	300,980
Croatian	3,35	96,65	7,384	212,935	220,319
Yoruba	2,34	97,66	763	31,860	32,623
Hausa	0,94	99,06	179	18,815	18,994
Igbo	1,15	98,85	126	10,858	10,984
Twi	1,56	98,44	34	2,146	2,180

The proportion of admin pages in relation to all articles across languages ranges between 1 to 17%. Smaller Wikis have a smaller percentage.

* as of August 2022

From a quick analysis using namespace 4 and 12 (Wikipedia and Help), we see that the proportion of admin pages in relation to all articles across languages ranges between 1 to 17%. Smaller Wikis have a smaller percentage.

Let's examine some of the characteristics of the admin pages to understand their value.
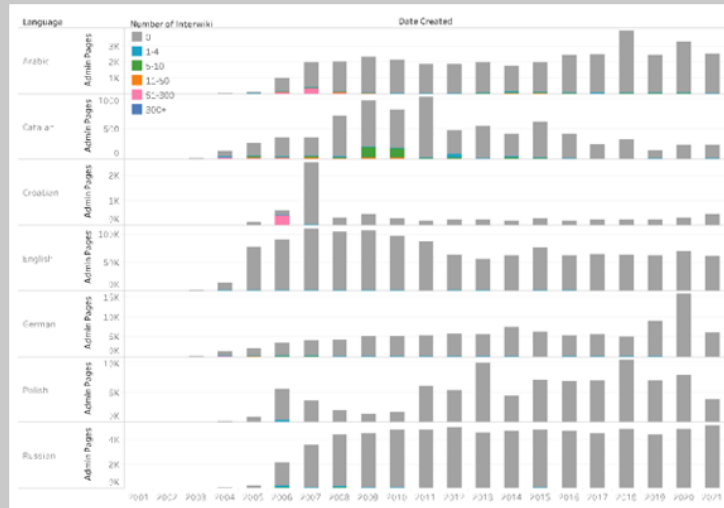
Creation of admin pages over time

There is a **constant creation** of admin pages over time. The majority are pages with no interwiki.

In this graph, we can see there is **constant creation** of admin pages over time. We see languages like Arabic, Catalan, German or Russian.

It is true that some languages have better years in the 2010s, but there is no pattern. Groups of Wikiprojects, Village Pump, help pages can be created by a rather small group of Wikipedians.

This does not mean that all these pages are valuable. In color, we see the pages by number of interwiki (grey means zero interwiki). We can see that pages with more interwiki were created in the ealy years.
They are pages that deserve to exist across languages.

Creation of *valuable* admin pages over time

The creation of valuable admin pages (> 1 interwiki link) over time across languages. Different number of admin pages = Content Gaps. **We have an opportunity for import-export across langs.**

This is the same graph as before, but we see only those pages with one or more interwiki links. We removed the grey ones with zero interwiki.

The pages with more interlanguage links in Orange and Pink, are in the early years.
If we pay attention to the number of pages, we see very different values across languages. English 1500, Catalan 200, Polish between 100 and 300…

This means that there is an opportunity for content gaps. Pages that are valuable across languages (with interwiki) but which do not necessarily exist in all languages.

We could actually learn more from other languages' admin pages. Export and Import.
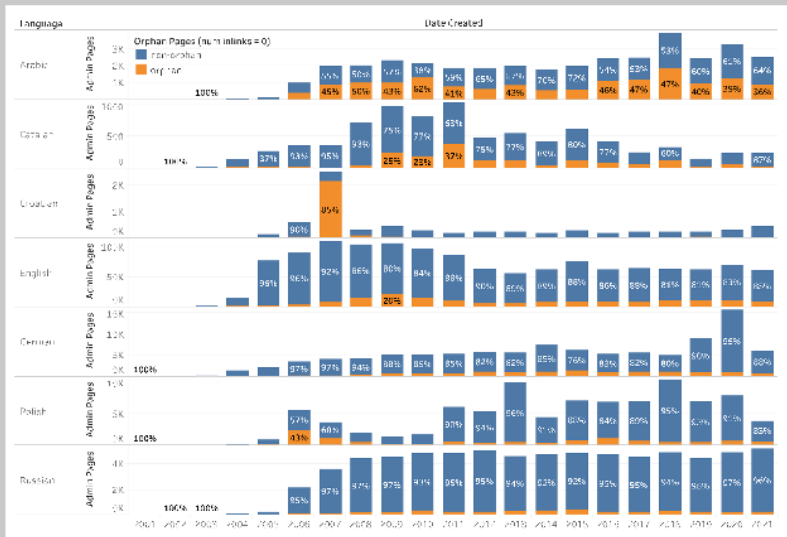
Creation of *orphan* admin pages over time

The creation of orphan (often non-valuable) admin pages over time across languages. Orphan = 0 inlinks.

In this graph, we also see the creation of pages over time, but colored by whether they are orphan pages or not. Orphan pages are those pages with zero inlinks (in orange).

If the interwiki were useful to see valuable pages, the number of inlinks also imply relevance.
Admin pages with no other pages pointing at them are les valuable.

There are orphan pages created over the years. We see pages created fifteen years ago and still orphan.

All these are potential pages to check and see if they are valuable. If so, link them from another page

Now we know about valuable pages (usually from the first years), and also the fact that there are many orphan page.
Let's see what is next.
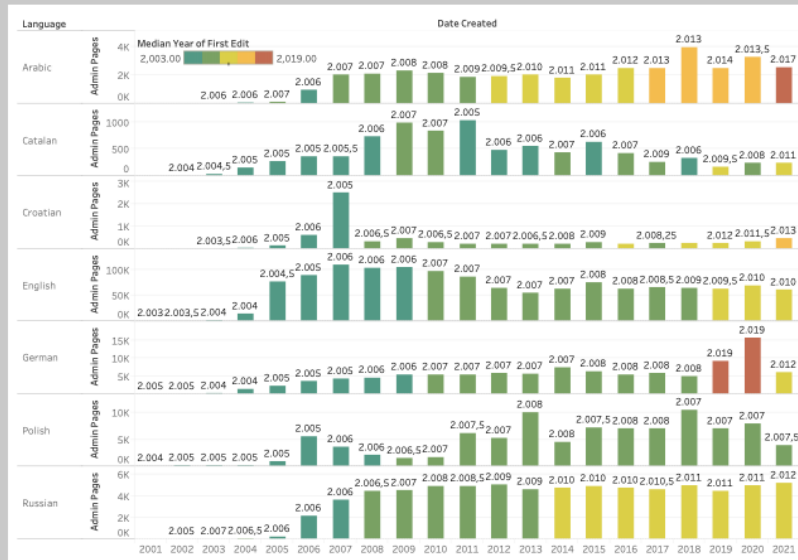
Creation of admin pages... by year of the first edit of their editors.
Created admin pages over time by the median year of the first edit of their editors (color).

In this graph, very similar to the previous ones, we see the number of
pages created colored by the median year of the first edit of the editors
who edited them.
It tells us about who creates the admin pages. Which generation of editors.

Greens go from 2003 until 2006 or 2009. We see that for some Wikipedias
like Catalan, English or Polish, greens are present in every year until
2021.
This means that even the new pages that are being created, they are
edited mostly by old editors.

Admin pages are generally not edited by newbies.

Percentage of admin pages by days since the last edit.
In most languages, the majority of admin pages has not been edited once in 2 years.

Finally, in this graph, we see in the 7 languages, the admin pages by when the last edit happened.

Orange is 1-2 years without editing. Red, Pink and purple, 5 until 20 years.

We see that a non-negligible part of the admin pages are not edited.

Something positive: in English Wikipedia (we see in green), a 40% of the admin pages have been edited once in the past 6 months. The green part is much smaller in other languages though.

Two conclusions stand out.

· **Content Gaps**
Very different number of valuable admin pages is available across languages, which means that there could be valuable content gaps, i.e., pages we could translate and help each other.

· **Under-edited Pages**
A small percentage of admin pages are edited in one year, which means that there are potential valuable non-edited pages that need more editing.

Editors of admin pages tend to be from previous generations. There is also an opportunity for inclusion.

Two conclusions stand out.

- Content Gaps. There are different number of valuable admin pages across languages, which means that there could be valuable "gaps". Pages we could translate and help each other.
- Under-edited Pages. Few pages are edited, which means that there are potential valuable non-edited pages that need more editing.

Editors of admin pages tend to be from previous generations. There is also an opportunity for inclusion. We could invite newcomers or newer editors to update these pages and take care of them.

Let's move to the second objective. 2. Classify the admin pages according to different types.

5.2 Assigning admin pages to types

First, choosing the types:

We look for categories:
- More interwiki
- More subcategories/pages

Serving different purposes.

6 main types

So, we now have a picture of the admin pages. But what types of admin pages do we have?

An easy way to start solving this problem is looking at the main categories within the category "Wikipedia administration" in English Wikipedia.

We looked at the subcategories with more interwiki (valuable across languages) and more subcategories and pages within them.

We identified 6 main types.

Main Types of Admin. Pages

The admin pages are in these two namespaces

Wikipedia (ns4)   Help (ns12)

Policies & Guidelines   Help Pages   Essays

Village Pump   Tools   Wikiprojects

https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics/Methodology
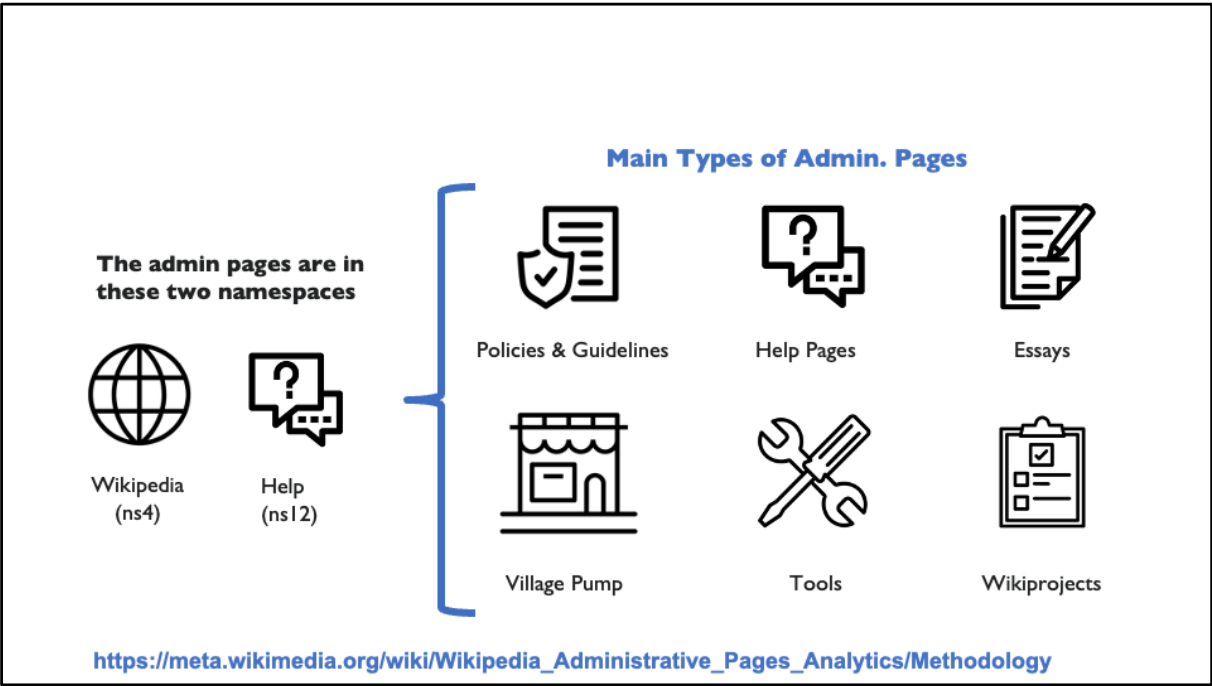
Policies & Guidelines, Help Pages, Essays, Village Pump, Tools, and Wikiprojects. All these are in namespace 4 and 12.

We assigned admin pages to types using the namespace, Wikidata properties and the category graph.

We selected the categories for the "category graph" that were the largest and with the most interwiki.

- Help Pages (ns12, wikidata property instance of: help page, category crawling)
- Policies & Guidelines (ns4, wikidata instance of, category crawling)
- Wikiprojects (ns4, wikidata instance of, category crawling)
- Essays (ns4, category crawling)
- Village Pump (ns4, category crawling)
- Tools (ns4, category crawling)
- Disclaimers (ns4, category crawling)
- Copyright (ns4, category crawling)

Often, one page belonged to multiple categories (e.g., essay and help page, policy and help page, etc.).

To collect and categorize admin pages in these 6 types, we used three strategies: namespaces, Wikidata properties, and category graph.

For example, Help Pages are identifiable using the namespace 12, the wikidata property instance of "help page" and the category "Help pages" and run it down.
We did this for the various types.

Often, one page belonged to multiple categories at the same time.
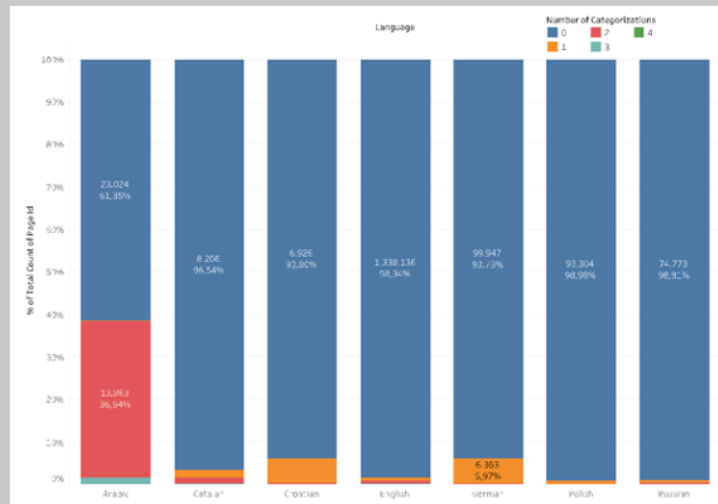
Most admin pages were not assigned.

Only a very small percentage of admin pages were identified through the methods exposed. We see that on average, it is below 5%.

The results showed the following.

Using the three approaches: namespaces, wikidata and category graph, we could only assign around 5% of the admin pages to the types. This is very very little.

First, we thought the approach is wrong, then, we understood that the majority of admin pages are not properly categorized in the category graph or in Wikidata.

These are the main types of admin pages. We also looked at the number of interwiki of the admin pages and saw that those with more interwiki tended to be more categorized.

It makes sense, pages with interwiki links are valuable in more than language. At the same time, editors take more care of categorizing them.

5.3 Applications

How can we use the analysis of admin pages to stimulate editing?
Helping editors find *valuable* pages.

Screening tools

https://en.wikipedia.org/wiki/Screening_(medicine)

At this point, we are at objective number 3. Create applications in order to support editing the admin pages.

We want to make something useful. We have understood what happens with admin pages, and their types.
How can we use the analysis of admin pages to stimulate editing?
Helping editors find *valuable* pages.

We need to make screening tools that allow us to do that.

What is a valuable and a non valuable admin page?

- *Completeness metrics:* Number of Bytes, Number of References, Number of Images, Number of Outlinks.
- *Relevance* metrics:* Number of Inlinks, Number of Interwiki, Number of Multilingual Sister Projects.
- *Popularity metrics:* Number of pageviews.

- **Activity**. Number of Edits, Number of Edits last month, Number of Anonymous Edits, Number of Bot Edits, Active months, Editing days.
- **Disagreement**. Number of Discussions, Number of Reverts.
- **Engagement**. Number of Editors.
- **Inclusion**. Number of admin editors, Number of Edits last month by newcomer 60d, Number of Edits last month by admin, Number of Edits last month by anonymous, Median year first edit, Median editors Edits.
- **Recency**. Days last 50 edits, Days last 5 edits, Days last edit, Date of last discussion, Date of last edit. Number of edits last month.
- **Regularity**. Percent active months, Percent active days, Max. Active Months Row, Max Inactive Months Row.
- **Creation**. Date of creation, Language of the first timestamp, Total months.

https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics/Metrics

What is a valuable and a non valuable admin page that the screening tool can find?

We need to rely on metrics for that.
Here, I list several dimensions, and some metrics that can represent them.

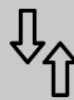For example, completeness is easily seen in the number of Bytes.
Relevance, the number of inlinks and interwiki links. Popularity, pageviews.

Other dimensions are: activity, disagreement, inclusion, recency, regularity, and creation.

The screening tools will need to allow us to use the dimensions and metrics in order to find pages.
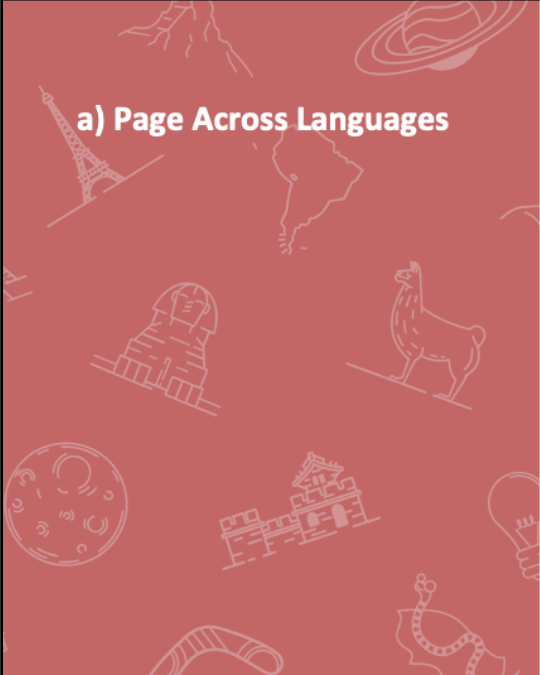
So, we initially created three screening tools. I must say that they are super super super experimental (alpha versión).

You can try them, but they may not work as expected. Because changes are happening every now and then.
So this is more an explanation of the concept.

a) Page Across Languages

⇓⇑

Introduce the page title and the metrics you want. Then you can compare the version of the page across languages.

Browse the results and sort the columns.

https://wapa.wmcloud.org/page_across_languages/

The first tool is called page across languages.

It is simple. It allows you to introduce a page title and the metrics you want. Then you can compare the version of the page across languages.

Browse the results and sort the columns.

Here, we introduce the page title, language or the Qitem.

Then the columns we want.

For example, these columns and the page "Neutral Point of View".

"Neutral Point of View" Page Characteristics Across Wikipedias

		languagecode	page_id	page_namespace	Title	Date of Creation	Number of Bytes	Number of Pageviews	Number of Interwiki	Number of Editors	Days Since The Last 5 Edits
		filter data									>1
×	☐	jawiki	270	4	中立的な観点	20030202063539	50534	1556	122	81	38
×	☐	ruwiki	8379	4	Нейтральная_точка_зрения	20040719194740	35779	3563	122	114	108
×	☐	frwiki	3243	4	Neutralité_de_point_de_vue	20020913135834	21658	2168	122	190	160
×	☐	itwiki	2987	4	Punto_di_vista_neutrale	20021218085625	6355	1491	122	82	174
×	☐	ptwiki	1034	4	Princípio_da_imparcialidade	20030801145040	13683	1033	122	104	308
×	☐	plwiki	3620956	4	Neutralny_punkt_widzenia	20020919131421	32189	583	122	80	333
×	☐	hrwiki	690614	4	Nepristrano_gledište	20050826094735	28858	180	122	28	371
×	☐	svwiki	1744	4	Neutral_synvinkel	20021125173138	21675	177	122	88	604
×	☐	hywiki	19693	4	Չեզոք_տեսանկյուն	20091025153748	2200	67	122	7	726
×	☐	dewiki	3542	4	Neutraler_Standpunkt	20011124160207	17144	906	122	290	867
×	☐	eswiki	2411	4	Punto_de_vista_neutral	20020723224232	11290	4095	122	105	882
×	☐	arwiki	12934	4	وجهة_النظر_المحايدة	20050615084124	7955	870	122	41	882
×	☐	cawiki	7170	4	Punt_de_vista_neutral	20040113235956	6370	120	122	24	3105
×	☐	rowiki	90	4	Punct_de_vedere_neutru	20030722001845	3674	171	122	15	3362
×	☐	igwiki	1948	4	Neutral_point_of_view	20070321023249	742	108	122	4	3439

Sortable Table

https://wapa.wmcloud.org/page_across_languages

The results show us Neutral Point of View Policy exists in only 121 Wikipedias out of the 320 Wikipedia language editions that exist today. This is something that should raise a red light.

In this table, we can also see that the page "Neutral Point of View" in Romanian and in Catalan has not had a five edits for more tan 3000 days.

b) Page Gaps

Search for admin pages in one language edition missing in another.

The second tool is called page gaps.

With it, we can search for admin pages in one language edition missing in another.

English Wikipedia Policies missing in Romanian Wikipedia

We choose:

- **Target Language (where gaps are)**
- **Source Language (where content is)**
- **Characteristics:**
  - **Admin page type**
  - **Admin namespaces**
  - **Retrieve by metric**

or search content (Wikidata, etc.).

- **Order the results by metric**

https://wapa.wmcloud.org/page_gaps/

We can select:

- Target Language (where gaps are): Romanian Wikipedia

- Source Language (where content is): English Wikipedia
- Characteristics:
  - Admin page type
  - Admin namespaces
  - Retrieve by metric → for example, number of inlinks.

or search content (Wikidata, etc.).

Then, a metric to order the results: for example, number of pageviews.

English Wikipedia Policies missing in Romanian Wikipedia

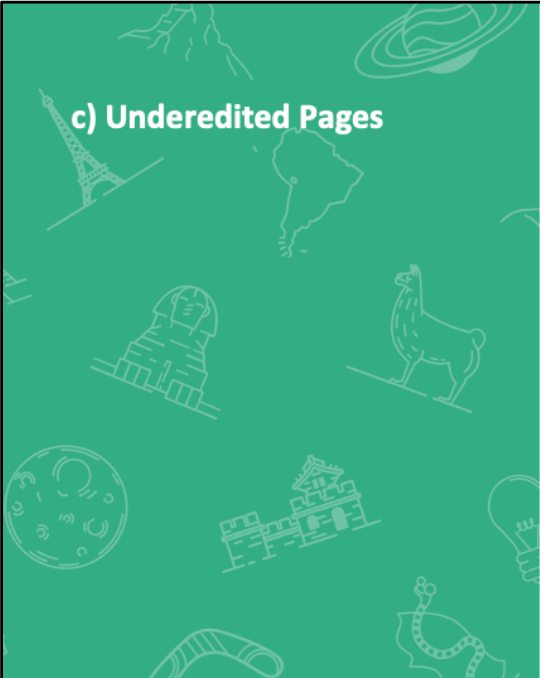| N°	Qitem	English Title	Edits	Days Since The Last Edit	Interwiki	Inlinks	Editors	Pageviews	Romanian Title
1	Q10973440	Wikipedia:Project namespace	806	2	23	24619	267	41170	
2	Q5854231	Wikipedia:Village pump (proposals)	139440	1	47	44331	8954	16606	
3	Q10973854	Wikipedia:Writing better articles	1316	19	39	537891	546	11547	
4	Q10416758	Wikipedia:Naming conventions (technical restrictions)	1068	58	17	1964	369	10613	
5	Q22964187	Wikipedia:Wikipedia is not a reliable source	552	4	7	1887	147	10476	
6	Q10966628	Wikipedia:Manual of Style/Lead section	2402	14	32	111062	720	9879	
7	Q10818201	Wikipedia:Administrators' noticeboard/Edit warring	232746	1	9	17902	16105	9618	
8	Q5617729	Wikipedia:Accuracy dispute	844	61	14	12637	285	9080	
9	Q4657739	Wikipedia:Non-free content	2097	71	54	1362768	702	8373	
10	Q10990487	Wikipedia:Copyright violations	355	103	24	84759	168	8060	
11	Q13360396	Wikipedia:Appealing a block	346	6	11	739799	158	7515	
12	Q5641607	Wikipedia:Talk page guidelines	2222	2	36	985235	782	7436	
13	Q11051015	Wikipedia:Content disclaimer	139	169	46	1690	58	7306	
14	Q10823887	Wikipedia:Wikimedia sister projects	568	39	30	16104	201	6743	
15	Q7114571	Wikipedia:Village pump (idea lab)	27062	0	11	10757	2554	6691	
16	Q11107170	Wikipedia:Reviewing pending changes	383	19	25	4344	168	6344	
17	Q11015023	Wikipedia:Legal disclaimer	70	169	51	677	45	6340	
18	Q15100248	Wikipedia:Categorizing redirects	449	73	3	2141033	135	6128	
19	Q20978602	Wikipedia:List of controversial issues	2161	24	5	5060	636	5925	
20	Q13429672	Wikipedia:Requests for undeletion	111606	2	20	66992	20493	5754	

Valuable one (?)

Table with the gaps

https://wapa.wmcloud.org/page_gaps/

Here, we have a table with the results.
English Policies missing in Romanian Wikipedia.

Here in the columns we see their number of Interwiki Links, Inlinks, Editors, and Pageviews, so that we can see their value.

Among them, for example, it strikes me the page "Wikipedia:Non-free content", which has more than one million inlinks and 8370 pageviews last month in English Wikipedia. Perhaps it should exist in Romanian Wikipedia.

The tool gives you ideas.

c) Underedited Pages

Find specific admin pages in Wikipedia language editions that are valuable, but might have not been edited enough, or for a long time.

And the third tool is called page gaps.

It allows you to find specific admin pages that may be valuable but have not been edited enough.

Polish Wikipedia: Help Pages with more interwiki/pageviews and less edited.

We choose:

- **Language edition**
- **Characteristics:**
  - **Admin page type**
  - **Admin namespaces**
- **Retrieve by metric (2 = Pareto Front)**
  - **Interwiki**
  - **Pageviews**
- **Select other columns**
  - **Date of creation, Bytes, Inlinks…**
- **Order the results by metric**
  - **e.g., "Days Since the Last Edit"**

https://wapa.wmcloud.org/underedited_pages/

So, let's take Polish Wikipedia as an example.

- We choose Language edition

- Characteristics:
  - Admin page type: We choose "Help Pages".
  - Admin namespaces. No need to select an admin namespace.

- Retrieve by metric
  - Interwiki
  - Pageviews

If we retrieve pages by two metrics, it computes the Pareto

35

front for these two metrics.
Pages in the frontier have a combination of values that no other page can have more of both at the same time.
Then, after the frontier pages, it lists the rest of pages with 0.5 weight for each metric.

- Select other columns
  - Date of creation, Bytes, Inlinks…

- Order the results by metric. Typically…
  - e.g., "Days Since the Last Edit"

Polish Wikipedia: Help Pages with more interwiki/pageviews and less edited.

The tool shows you some of the pages represented in a scatterplot.

Polish Wikipedia: Help Pages with more interwiki/pageviews and less edited.

N°	Qitem	Page Title	Bytes	Days Since The Last Edit	Interwiki	Pageviews	Date of Creation	Edits	Editors	Inlinks	Pareto Front
1	Q10640405	WP:Opisy licencji grafiki	9.4k	861	52	28	2004-09-06	196	38	5780	Yes
2	Q13921062	WP:VisualEditor/Dlaczego	3.8k	633	8	15	2013-07-15	12	5	14	Yes
3	Q56723771	WP:PetScan	1.3k	626	14	25	2011-08-30	20	6	35	Yes
4	Q3396208	WP:Narzędzia/HotCat	1.9k	592	96	49	2008-04-15	40	11	58	Yes
5	Q3223292	WP:Szablony/Dyskusja wikipedysty	11.6k	455	36	22	2005-03-06	181	76	223	Yes
6	Q13666672	WP:VisualEditor/Nawigacja	1.2k	412	35	16	2013-07-10	25	7	29	Yes
7	Q10634368	WP:Najczęstsze nieporozumienia	8.7k	306	39	60	2002-11-25	93	36	25286	Yes
8	Q4048867	WP:Boty	16.8k	8	185	97	2003-03-04	468	152	649	Yes
9	Q6584911	WP:Narzędzia	22.2k	2	41	84	2006-08-01	158	54	1317	Yes
10	Q4167410	WP:Strona ujednoznaczniająca	12.6k	160	113	4925	2002-11-23	711	141	190878	No
11	Q4582194	WP:Kawiarenka/Kwestie techniczne	85.9k	1	76	3275	2006-04-14	41846	1101	3440	No
12	Q11168190	WP:VisualEditor	8.3k	87	145	597	2013-04-27	173	50	0	No
13	Q4592334	WP:Strona użytkownika	14.0k	265	97	571	2007-02-24	175	59	0	No
14	Q3186256	WP:Nazewnictwo artykułów	9.2k	433	101	82	2013-05-18	58	23	20201	No
15	Q4063270	WP:AutoWikiBrowser	2.6k	361	102	23	2006-09-15	58	20	63	No
16	Q4582561	WP:Zadania dla botów	21.1k	1	52	1549	2005-05-12	20134	596	649	No
17	Q4664077	WP:Nazwa użytkownika	4.8k	207	81	557	2005-11-27	196	79	0	No
18	Q4615128	WP:Boty/Zgłoszenia	2.9k	623	91	25	2006-09-30	418	88	322	No
19	Q3916099	WP:Śmiało modyfikuj strony	2.1k	2	84	205	2002-11-28	244	93	234302	No
20	Q28521	WP:Łasica	3.4k	450	80	24	2007-02-11	44	19	205	No

Valuable one (?)

Table sorted by days since the last edit

And the table. Here we see the Polish Wikipedia Help Pages with more interwiki/pageviews and les edited.

The first page is called Wikipedia:Opisy licencji grafiki. Please, forgive me. It might something like "Graphic license descriptions."

It exists in 52 languages, it has more than 5000 inlinks and has not been edited for more than 2 years.
Maybe it would be time to take a look at this page.

In general, I see good maintenance in Polish Wikipedia compared to other languages.

This third tool, "Underedited Pages", it allows you seeing valuable pages that need maintenance in the jungle of admin pages.

Conclusions

Conclusions

I can see that the small Wikipedias can easily grow their *relevant* admin pages. There are many gaps and it is surprising that the policy Neutral Point of View only exists in 121 Wikipedia language editions out of the 320. Notability exists in 96. We have work to do. We need to support small Wikipedias and tell them to create the basics. Let's fill the admin pages gaps!

At the same time, we also need to keep updated the most relevant pages. Let's reign in chaos. Most importantly, let's invite newcomers. Seeing that new created pages are also edited by veteran editors should makes us reflect on whose Wikipedia it is.

- Admin pages are essential; they should be a bridge between Wikipedians, between generations.

- Visualizations can help in getting a glimpse of how well we are doing it.

- Analytical tools do not necessarily tell you to delete, but they can help you focus.

What can you do as **CEE communities**?

Editors, keep on editing. Affiliates: Please, create programs and get organized around admin pages.

We drew the ideas for the tools from conversations with fellow Wikimedians to identify "use cases" to generate tools.

Please, talk out your ideas: we need to identify and polish the use cases. Maybe there are better tools to make. All feedback is valuable!

Looking forward to Wikimania 2023.

What can you do as CEE communities?

Advice? I cannot give much. But I have to say that affiliates have the capacity to organize work and push harder.

You can also give more feedback on this project.

We drew the ideas for the tools from conversations with fellow Wikimedians to identify "use cases" to generate tools.

All feedback is super valuable.

What can we do as **researchers**?

- Improving the dashboards so that they encompass more use cases.
- Make available all the datasets and encourage research on admin pages.
- Bridges with product in order to help in the use of data.
- Other research lines: aspects related to readability.

If you have any **question**, you can e-mail me at
marcmiquel@gmail.com or Username:marcmiquel
or leave me a message at the project page.

https://meta.wikimedia.org/wiki/Wikipedia_Administrative_Pages_Analytics

I don't think these will be the ultimate tools, but I believe that these proofs of concept are essential.

We need to improve the dashboards, make them more usable and encompass more use cases.

Wikipedia Admin Page Analytics is the first project to analyze admin pages. We can make the datasets available and encourage more research and applications.
In the future, other research lines can give place to more tools, for example, measuring the readability of the admin pages.

We need to open this avenue of work on admin pages because they really really matter.

If you have any question, please e-mail me.

Finally, I want to close with the two golden mottos of the project.

1. Admin pages with great power, come with great maintenance.

2. If it is everybody's Wikipedia, it must be everybody's admin pages.

Let's think about them. Thank you very much.

Thank you very much!

Marc Miquel i Ribé, PhD
User:marcmiquel / Catalan Wikipedia
marcmiquel@gmail.com