

Looking for possible new articles: how to use PetScan



WIKIMEDIA
EDUCATION

Who are we?

Meghan Dowling

- Wiki: Dowlinme
- ADAPT Centre, DCU

Kepa Sarasola

- Wiki: KSarasola
- IXA group, EHU/UPV

Ana Zelaia Jauregi

- Wiki: Ana Zelaia Jauregi
- EHU/UPV

Aitzol Astigarraga

- Wiki: Aitzolete
- Basque Wikipedia
-

What is Petscan?

An external **tool** that searches an article category (and its subcategories) according to specified criteria to find articles, stubs, images, and categories.



Finding new articles

Petscan

+

Massviews

Find articles that are present on 1+ Wikipedias & not present in target language Wikipedia

Rank these articles by number of pageviews



STEP 1:

<https://petscan.wmflabs.org>



Categories

Page properties

Templates&links

Other sources

Wikidata

Output

Language

en

[Commons](#) [Wikispecies](#) [Wikidata](#)

Project

wikipedia

Depth

2



(depth 0 means no subcategories; appending '|' and a number to a category in the fields below will override it for that category)

Categories

economy

Combination

Subset Union

Negative categories

Do it!

Link to a pre-filled form for *the query you just ran* [with](#) and [without](#) auto-run. PSID is [7265158](#). Find images for [Wikidata](#).

STEP 2:

Category

Language

Project

Depth

Categories

en

wikipedia

2

(depth 0 means no subcategories, a per category)

economy

2

economy

Choose a depth and category

Do it!

Link to a pre-filled form for *the query you just ran* with and without auto-run. PSID is **7265158**. Find images for Wikidata.

STEP 3:

Click on the
'Wikidata' tab

https://petscan.wmflabs.org

PetScan Manual Issues English Examples List

Categories Page properties Templates&links Other sources **Wikidata** Output

Pages with items Do not add Wikidata items Add items, where available Only pages with item
 Only pages without item

Label language Defaults to interface language

Uses items/props Q123,P456,... Any All None

Has no statements Has no sitelinks

Site links *Note: The options below are filters, not generators!*

Has all of these site links	Has any of these site links	Has none of these site links
<input type="text" value="enwiki, frwiki, ruwiki, cawiki"/>	Site names, one per row (e.g. enwiki, dewikisource)	<input type="text" value="qawiki"/>
Minimum number of site links		Maximum number of site links

Labels etc. *Note: The options below will be used as a generator, if and only if no other generator is used!*

Has all of these labels	Has any of these labels	Has none of these labels
One per line, SQL LIKE syntax (e.g., "Bob %")	One per line, SQL LIKE syntax (e.g., "Bob %")	One per line, SQL LIKE syntax (e.g., "Bob %")
<input checked="" type="checkbox"/> Label <input type="checkbox"/> Alias <input type="checkbox"/> Desc.	<input checked="" type="checkbox"/> Label <input type="checkbox"/> Alias <input type="checkbox"/> Desc.	<input checked="" type="checkbox"/> Label <input type="checkbox"/> Alias <input type="checkbox"/> Desc.
Languages; en,de...	Languages; en,de...	Languages; en,de...

Link to a pre-filled form for the query you just ran with and without auto-run. PSID is 7265158. Find images for

STEP 4:

Choose which sites to filter by

Site links

Note: The options below are *filters*, not generators!

Has all of these site links

enwiki
frwiki
ruwiki
cawiki

Has any of these site links

Site names, one per row
(e.g. enwiki, dewikisource)

Has none of these site links

gawiki



STEP 5:

Click on the
'Output' tab

The screenshot shows the PetScan website interface. At the top, there is a navigation bar with 'PetScan', 'Manual', and 'Issues' links. A language dropdown menu is set to 'English'. On the right side of the navigation bar, there are buttons for 'Examples' and 'List'. Below the navigation bar, there are several tabs: 'Categories', 'Page properties', 'Templates&links', 'Other sources', 'Wikidata', and 'Output'. The 'Output' tab is highlighted with a blue circle, and a blue arrow points from a text box to it. The main content area contains various settings for the search, including radio buttons for output format (HTML, CSV, TSV, Wiki, JSON, PagePile), sorting options (not at all, by title, by namespace, title, by size, by date, by incoming links, by file size, by upload date, randomly), and sort order (ascending, descending). There are also checkboxes for extended data, file usage data, main namespace only, generate thumbnails, image, coordinates, default sort, and disambiguation. A 'Regexp filter' section has a text input field and a checkbox for 'Only show titles/labels matching the regular expression'. A 'Redlinks' section has checkboxes for 'Show redlinks', 'Remove redlinks linked from templates', and 'Show only redlinks to main (article) namespace', along with a 'Minimum redlink count' input field set to 1. A 'Limit' section has a 'Max results (10K in browser)' input field. At the bottom, there is a 'Do it!' button and a link to a pre-filled form for the query.

https://petscan.wmflabs.org

PetScan Manual Issues English

Examples List

Categories Page properties Templates&links Other sources Wikidata Output

HTML CSV TSV Wiki JSON PagePile

CatScan QuickIntersection Sparse (prefixed page titles instead of objects)

Pretty (not machine-readable)

Sort not at all (outputs by page ID = upload date for files) by title by namespace, title

by size by date by incoming links (ns0) by file size by upload date randomly

Sort order ascending descending

Extended data for files File usage data Main namespace only

Generate thumbnails in Wiki output

metadata Image Coordinates Default sort Disambiguation

Regexp filter Only show titles/labels matching the regular expression

Redlinks Show redlinks Remove redlinks linked from templates

Show only redlinks to main (article) namespace

Minimum redlink count 1

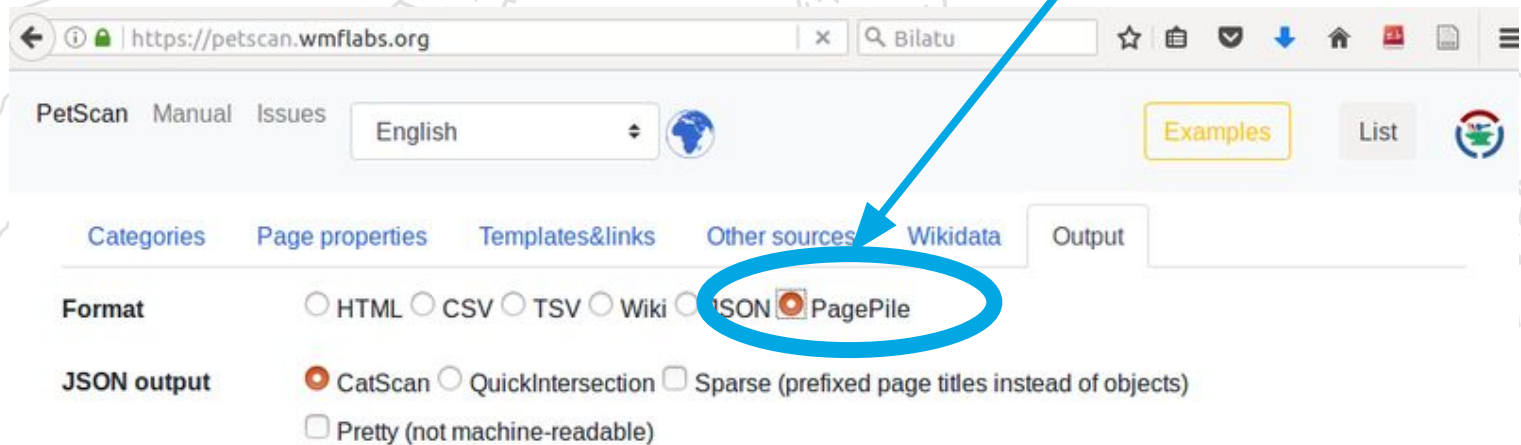
Limit Max results (10K in browser)

Do it!

Link to a pre-filled form for *the query you just ran with* and *without* auto-run. PSID is [7265158](#). Find images for [Wikidata](#).

STEP 6:

Choose 'Pagepile'
as the format



The screenshot shows the PetScan web interface. The browser address bar displays `https://petscan.wmflabs.org` and the search term `Bilatu`. The page title is `PetScan`. The language is set to `English`. There are buttons for `Examples` and `List`. The `Other sources` tab is selected, and the `Format` section is visible. The `PagePile` radio button is selected and circled in blue. The `JSON output` section has `CatScan` selected.

Categories Page properties Templates&links Other sources Wikidata Output

Format HTML CSV TSV Wiki JSON PagePile

JSON output CatScan QuickIntersection Sparse (prefixed page titles instead of objects)
 Pretty (not machine-readable)

STEP 7:

Limit

Do it!

Minimum re

Max results

Link to a pre-
Wikidata.

RESULT:

[File 19847](#): 506 pages on enwiki. Other formats: [Plain text](#), [wikitext](#), [JSON](#), [J](#)

Use this pagepile in

1-100 | [101-200](#) | [End](#)

1. [Anarcho-capitalism](#)
2. [Austrian School](#)
3. [Benelux](#)
4. [Barter](#)
5. [Balance of trade](#)
6. [Board of directors](#)
7. [Colony](#)
8. [Corporation](#)
9. [Cash register](#)
10. [Consumer](#)
11. [Planned economy](#)
12. [Data warehouse](#)
13. [Dollar](#)
14. [Denarius](#)
15. [Dutch West India Company](#)
16. [Distributism](#)
17. [Division of labour](#)
18. [Enterprise resource planning](#)
19. [Egalitarianism](#)
20. [Finance](#)

RESULT:

File **19847**: 506 pages on enwiki. Other formats: [Plain text](#), [wikitext](#), [JSON](#), [J](#)

Use this page in

1-100 | [101-200](#) | [End](#)

1. [Anarcho-capitalism](#)
2. [Austrian School](#)
3. [Benelux](#)
4. [Barter](#)
5. [Balance of trade](#)
6. [Board of directors](#)
7. [Colony](#)
8. [Corporation](#)
9. [Cash register](#)
10. [Consumer](#)
11. [Planned economy](#)
12. [Data warehouse](#)
13. [Dollar](#)
14. [Denarius](#)
15. [Dutch West India Company](#)
16. [Distributism](#)
17. [Division of labour](#)
18. [Enterprise resource planning](#)
19. [Egalitarianism](#)
20. [Finance](#)

We'll use this in the next step



STEP 8:

[https://tools.wmflabs.org/
massviews/](https://tools.wmflabs.org/massviews/)



Massviews Analysis

Import a list of pages and compare the pageviews

Dates

2018-12-28 - 2019-01-17

Platform

All

Agent ?

User

Source ?

Page Pile ▾ 19847

Submit

Get the pageviews of pages in a [PagePile](#)



STEP 9:

Choose Page Pile as the source and enter your pagepile number

https://tools.wmflabs.org/massviews/

Pageviews Langviews Topviews Siteviews Massviews More

Massviews Analysis

Import a list of pages and compare the pageviews

Dates
2018-12-28 - 2019-01-17

Platform
All

Source
Page Pile 19847

Get the pageviews of pages in a PagePile

https://tools.wmflabs.org/

File 19847: 506 page

Use this pagepile in

Submit



RESULT:

Massviews Analysis

Import a list of pages and compare the pageviews

Page Pile 19847 2018-12-28 - 2019-01-17

[List](#)[Chart](#)[Permalink](#)[Download ▾](#)

#	Page title	Pageviews ↓	Daily average
Totals	506 pages	6,778,117	322,767
1	Human Development Index	157,304	7,491 / day
2	List of countries by GDP (PPP) per capita	154,443	7,354 / day
3	East India Company	106,806	5,086 / day
4	World Bank	106,100	5,052 / day
5	Marxism	87,037	4,145 / day
6	Six Sigma	83,402	3,972 / day
7	Libertarianism	67,193	3,200 / day
8	Airbnb	63,703	3,033 / day
9	Great Pacific garbage patch	61,005	2,905 / day

Thank you!
Go raibh maith agat!
Eskerrik asko!



WIKIMEDIA
EDUCATION