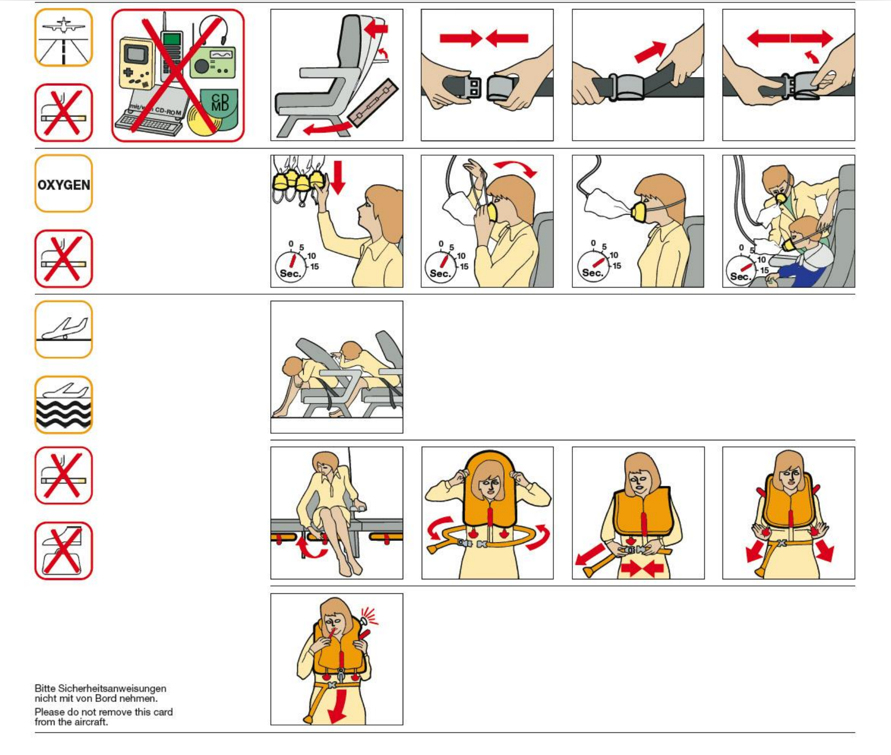
Visual Enrichment of Collaborative Knowledge Bases

Miriam Redi
& Special Guest Magnus!

WIKIMEDIA
FOUNDATION

We Love **Images**

Photo by Dominik Scythe on Unsplash



Bitte Sicherheitsanweisungen nicht mit von Bord nehmen.

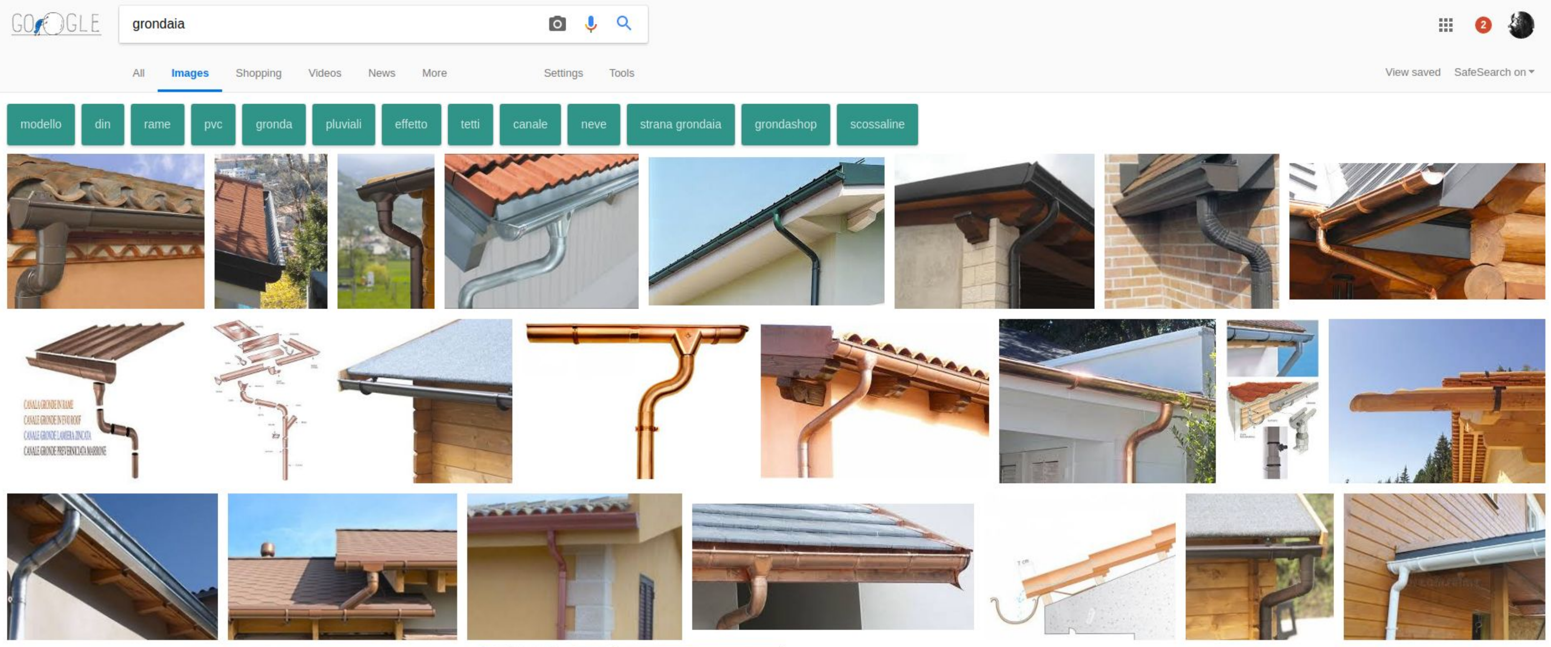
Please do not remove this card from the aircraft.

Visual **Language**

perception, comprehension and production of visible signs

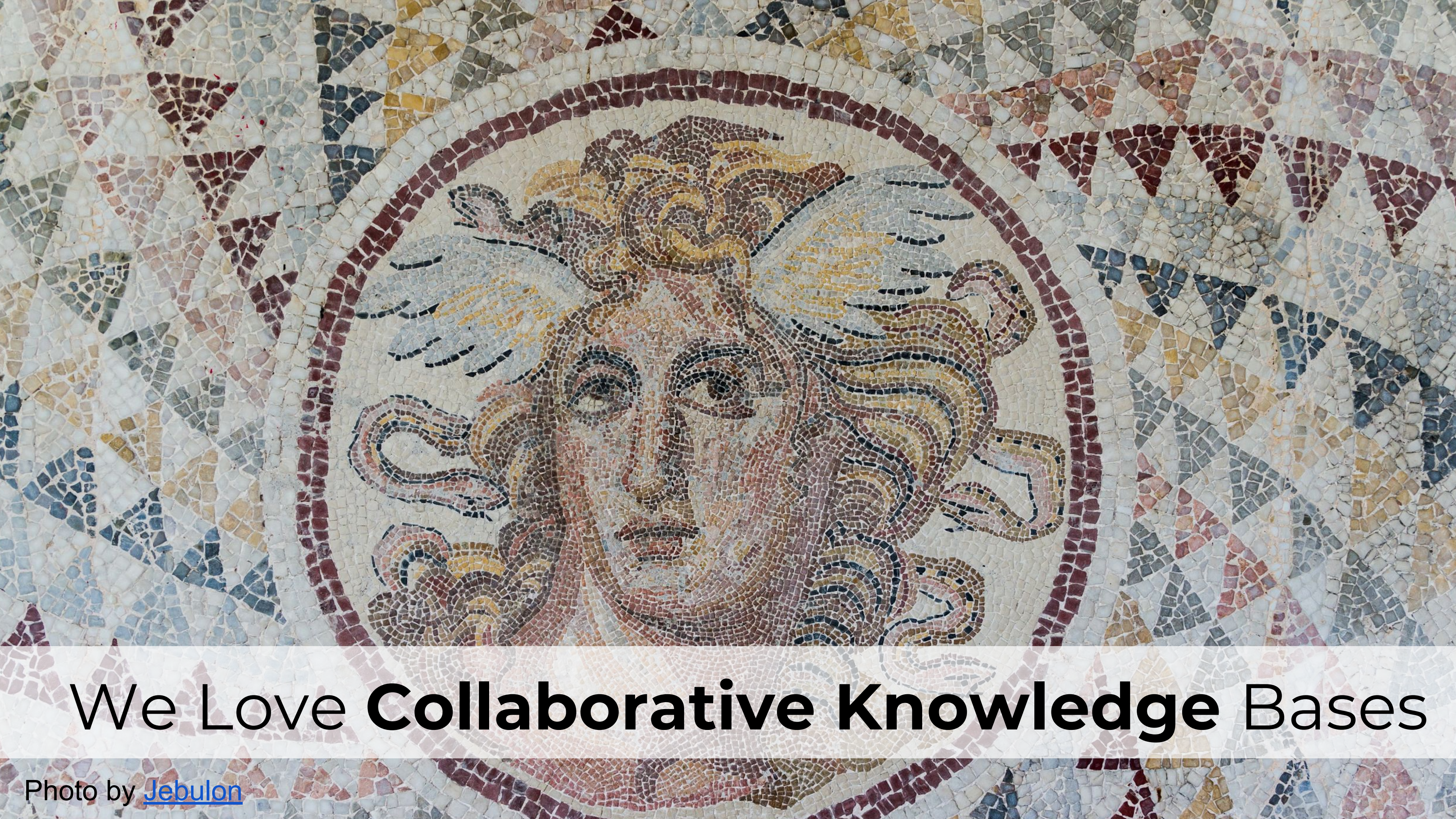
Images allow to explain, enrich and complement knowledge without language barriers[1]

[1] Van Hook, S.R. (2011, 11 April). Modes and models for transcending cultural differences in international classrooms. Journal of Research in International Education, 10(1), 5-27.



grondaia

“Rain Gutter”

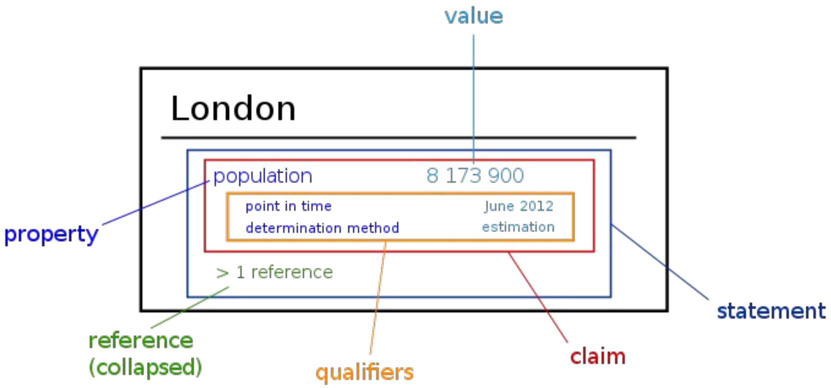


We Love **Collaborative Knowledge** Bases

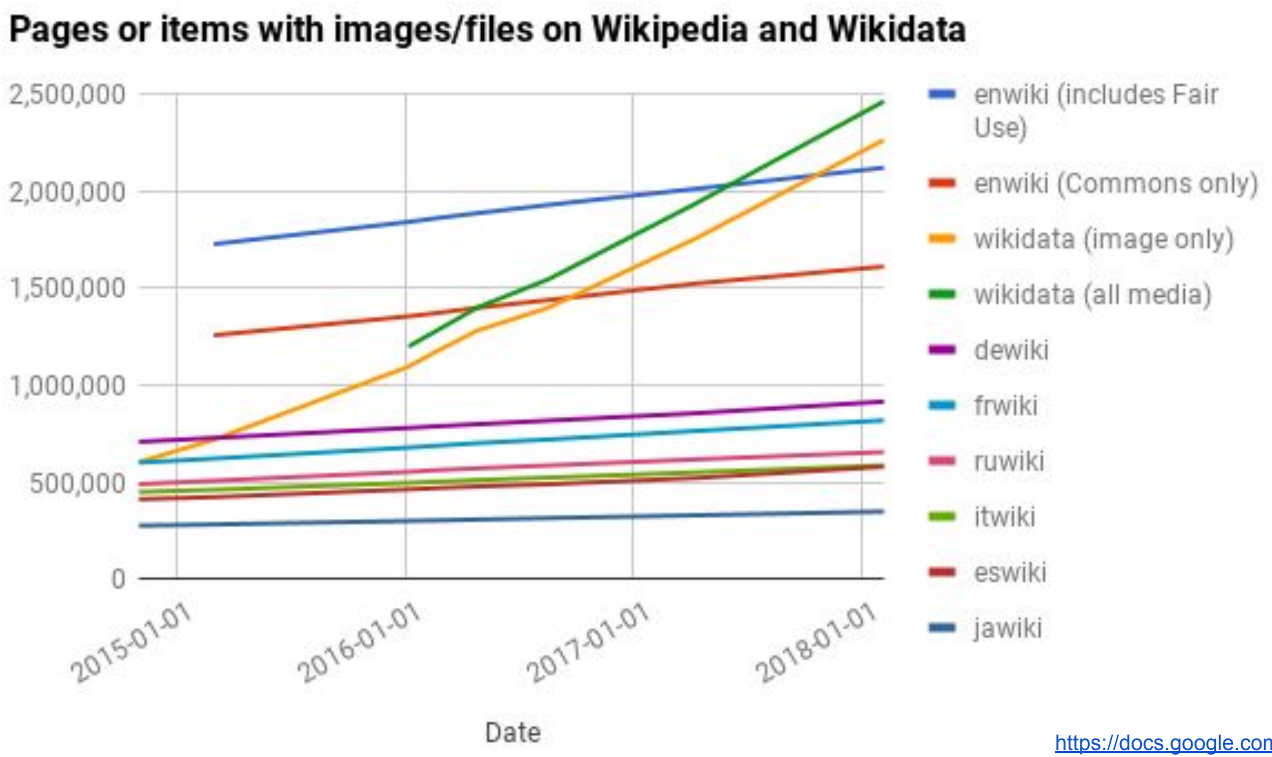
Photo by Jebulon

Wikidata

Wikidata is an international and thus multilingual project. While English is the default interface language, the project is intended to be used by, and useful for, users of every language with MediaWiki internationalization support.



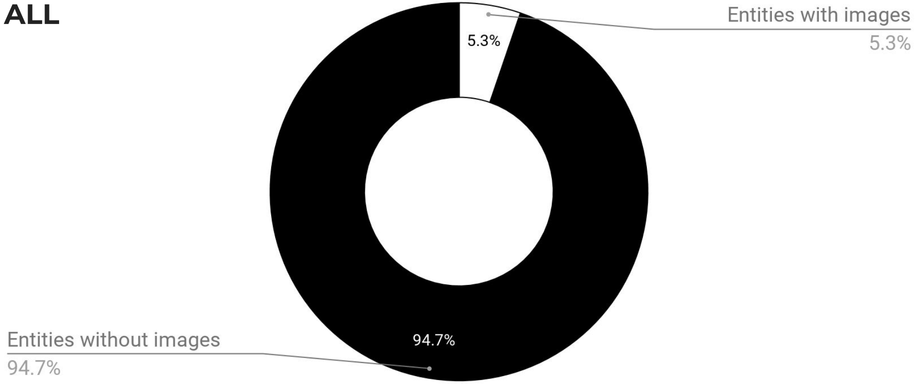
Wikidata already has the most images



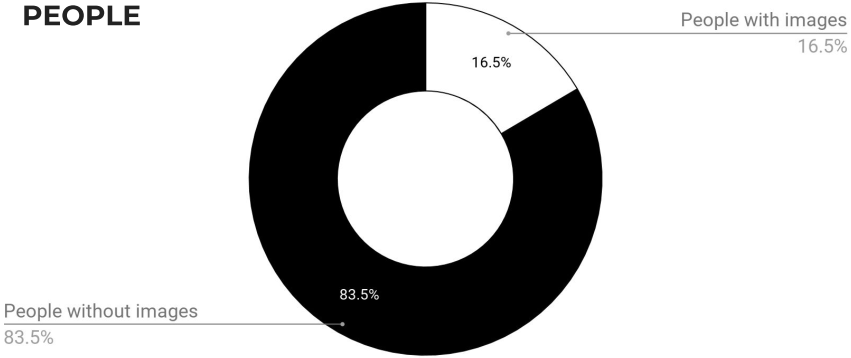
https://docs.google.com/spreadsheets/d/1G2P89WW_GAkqShvaWdhkiCN66WZB0synLCV9X0h6-Ig/edit?usp=sharing

But Wikidata is still missing images!

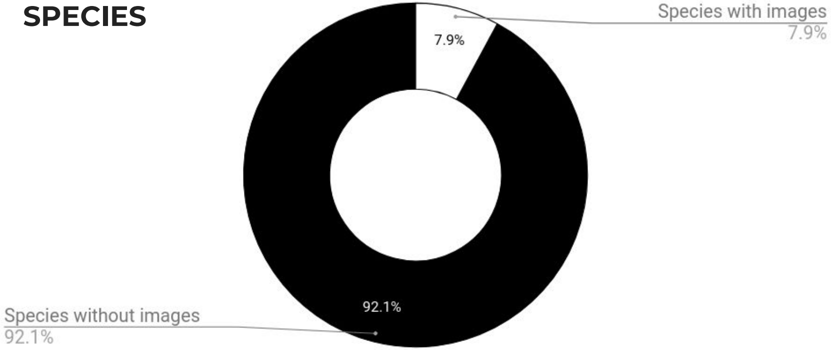
ALL



PEOPLE



SPECIES



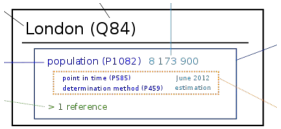
Smart Tools for **Wikidata Visual Enrichment**

Photo by Jess Watters on Unsplash

Visual Enrichment: Problem

- Worst-case scenario: users willing to add images to Wikidata might need to manually search for the right image using different tools from over **40M free-licensed commons images**
- We can help reducing the search space by
  - **pulling images from reliable sources**
  - **using content-based analysis**

Visual Enrichment: Pipeline



London (Q84)

population (P1082) 8 173 900

point in time (P585) June 2012

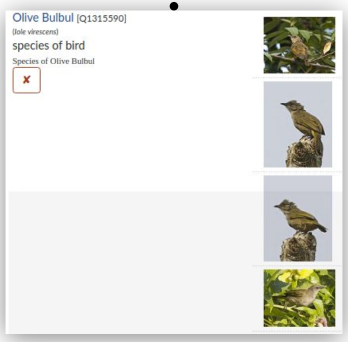
determination method (P459) estimation

> 1 reference

Item without
P18 - ‘Has image’

Discovering related
images from different
sources

Ranking images
according to
Relevance and Quality



Olive Bulbul [Q1315590]

(*Iole virescens*)

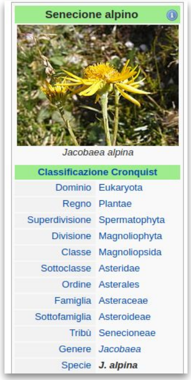
species of bird

Species of Olive Bulbul

Manual Selection and
Evaluation

1. **Finding Free-Licensed images**

From: linked pages, Commons, Flickr

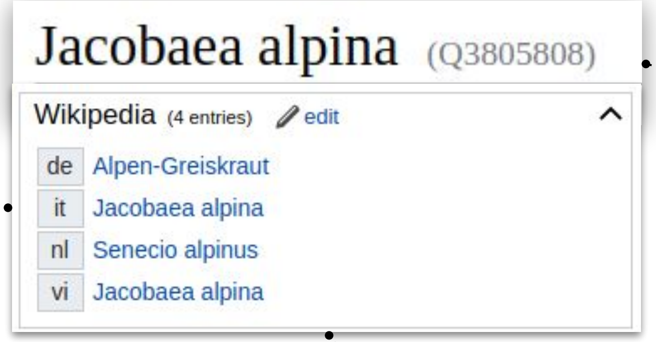


Senecione alpino

Jacobaea alpina

Classificazione Cronquist

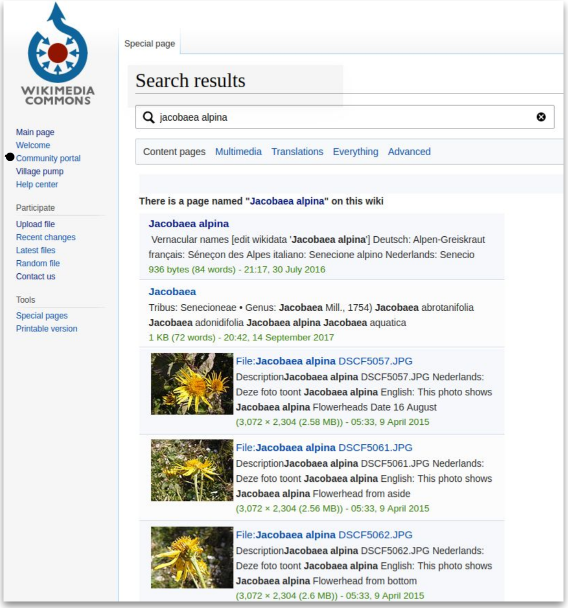
Dominio	Eukaryota
Regno	Plantae
Superdivisione	Spermatophyta
Divisione	Magnoliophyta
Classe	Magnoliopsida
Sottoclasse	Asteridae
Ordine	Asterales
Famiglia	Asteraceae
Sottofamiglia	Asteroideae
Tribù	Senecioneae
Genere	*Jacobaea*
Specie	***J. alpina***



Jacobaea alpina (Q3805808)

Wikipedia (4 entries) edit

de	Alpen-Greiskraut
it	Jacobaea alpina
nl	Senecio alpinus
vi	Jacobaea alpina



Search results

jacobaea alpina

Content pages Multimedia Translations Everything Advanced

There is a page named "Jacobaea alpina" on this wiki

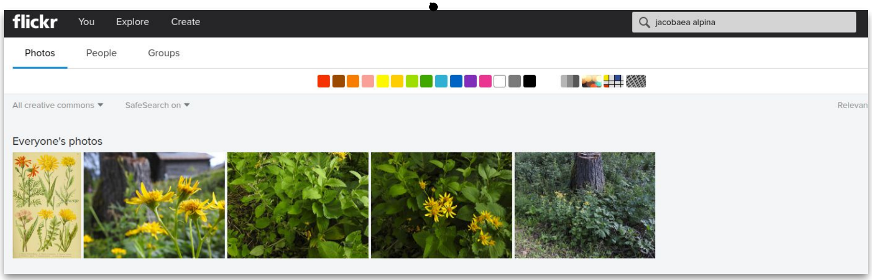
Jacobaea alpina
Vernacular names [edit wikidata '**Jacobaea alpina**'] Deutsch: Alpen-Greiskraut français: Séneçon des Alpes italiano: Senecione alpino Nederlands: Senecio
936 bytes (84 words) - 21:17, 30 July 2016

Jacobaea
Tribus: Senecioneae • Genus: **Jacobaea** Mill., 1754) **Jacobaea** abrotanifolia **Jacobaea** adonidifolia **Jacobaea alpina Jacobaea** aquatica
1 KB (72 words) - 20:42, 14 September 2017

File:**Jacobaea alpina** DSCF5057.JPG
Description**Jacobaea alpina** DSCF5057.JPG Nederlands: Deze foto toont **Jacobaea alpina** English: This photo shows **Jacobaea alpina** Flowerheads Date 16 August
(3,072 × 2,304 (2.58 MB)) - 05:33, 9 April 2015

File:**Jacobaea alpina** DSCF5061.JPG
Description**Jacobaea alpina** DSCF5061.JPG Nederlands: Deze foto toont **Jacobaea alpina** English: This photo shows **Jacobaea alpina** Flowerhead from aside
(3,072 × 2,304 (2.56 MB)) - 05:33, 9 April 2015

File:**Jacobaea alpina** DSCF5062.JPG
Description**Jacobaea alpina** DSCF5062.JPG Nederlands: Deze foto toont **Jacobaea alpina** English: This photo shows **Jacobaea alpina** Flowerhead from bottom
(3,072 × 2,304 (2.6 MB)) - 05:33, 9 April 2015



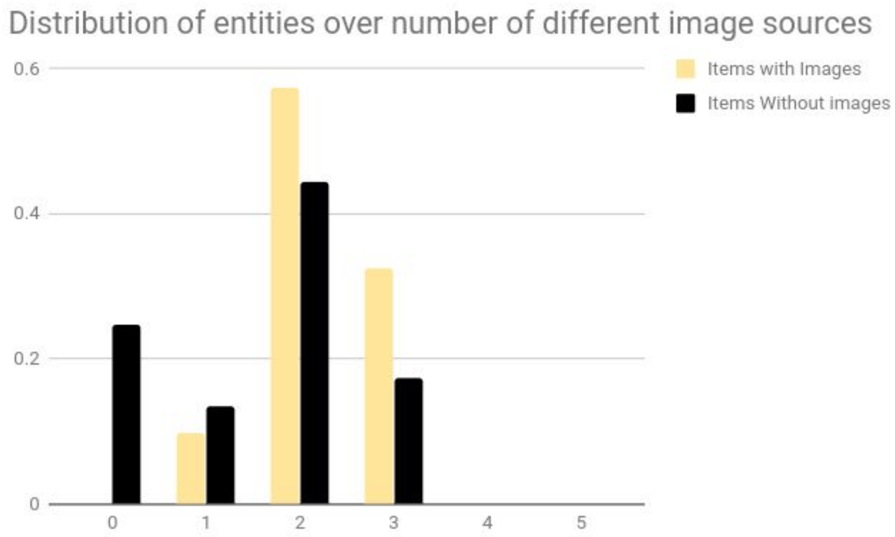
jacobaea alpina

All creative commons ▾ SafeSearch on ▾

Everyone's photos

1. **Finding Free-Licensed images**

Around 75% of people entities have images linked to them



There are so many of them.. How to reduce the amount of images users should select from?

2. **Ranking images**

Not all image candidates gathered are good candidates for the item

RELEVANCE AND QUALITY

RELEVANCE: does the images depict the entity?

QUALITY: is the image of high photographic quality?

2. **Ranking images** (relevance)

RELEVANCE: does the images depict the entity?

- Candidates are already biased towards relevance
- **Temporary solution: match image name/desc with item label** (plus face detection)
- **Content-based techniques** hard to use due to the singularity of each entity

ONGOING: semantic distance between item description and objects recognized (word2vec)

2. **Ranking images** (quality)

QUALITY: is the image of high photographic quality?

Not all relevant images are actually ‘good’ images

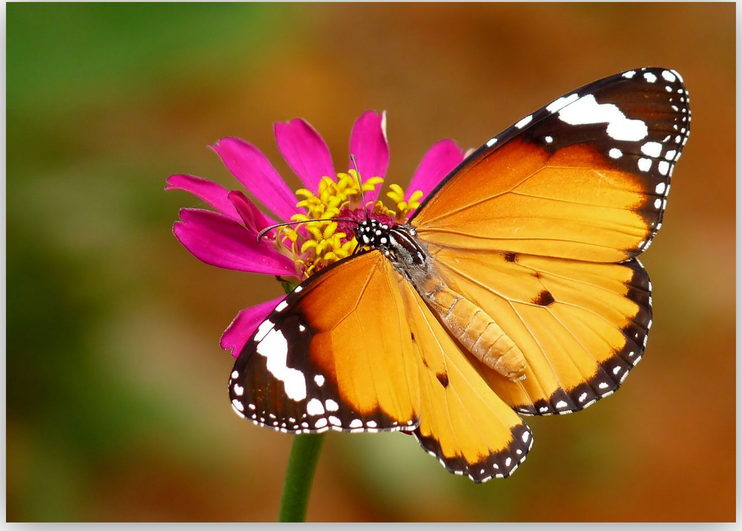


Photo: Jee & Rani Nature Photography on Commons

High Quality

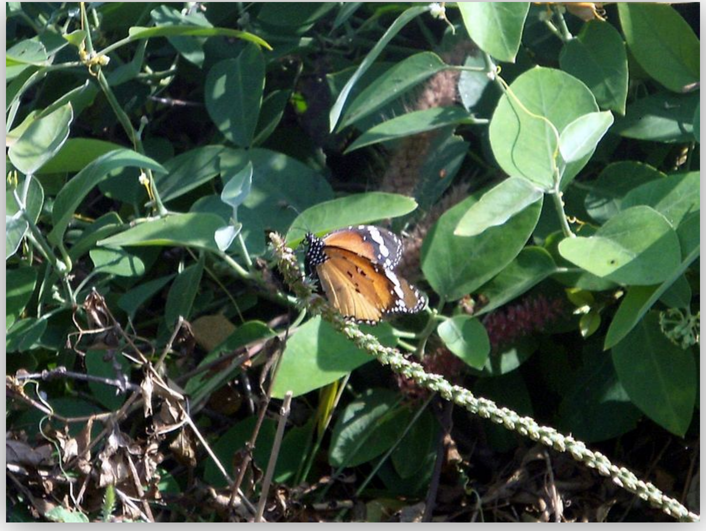


Photo: Vinayaraj on Commons

Lower Quality

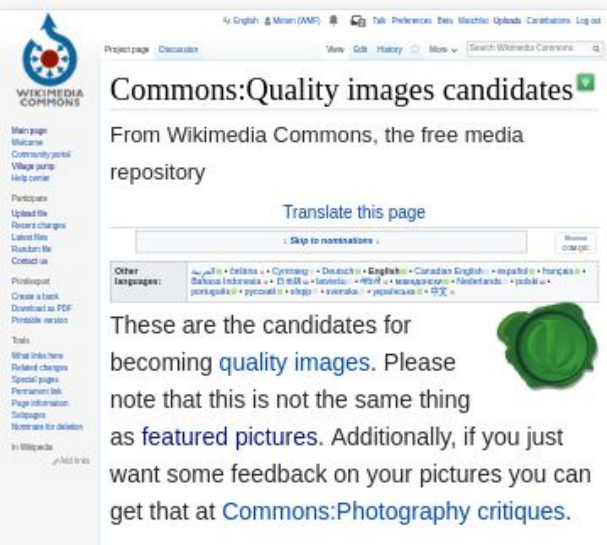
2. **Ranking images** (quality)

Solution: **COMPUTATIONAL AESTHETICS**

- Branch of Computer Vision
- Exploits visual information to detect quality
- Uses **Supervised Learning** Techniques
- Needs images **annotated** as being ‘High’ or ‘Low Quality’

2. **Ranking images** (quality)

DATA: exploiting the richness of the Commons



Commons:Quality images candidates

From Wikimedia Commons, the free media repository

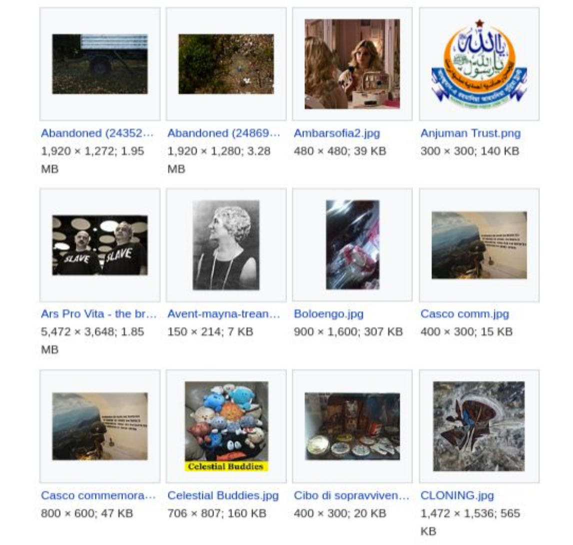
Translate this page

↓ Skip to nominations ↓

Other languages: العربية • čeština • Cymraeg • Deutsch • English • Canadian English • español • français • Bahasa Indonesia • 日本語 • latviešu • हिन्दी • македонски • Nederlands • polski • português • русский • shqip • svenska • українська • 中文

These are the candidates for becoming quality images. Please note that this is not the same thing as featured pictures. Additionally, if you just want some feedback on your pictures you can get that at Commons:Photography critiques.

High Quality:
160K Quality Commons



Abandoned (24352…
1,920 × 1,272; 1.95 MB

Abandoned (24869…
1,920 × 1,280; 3.28 MB

Ambarsofia2.jpg
480 × 480; 39 KB

Anjuman Trust.png
300 × 300; 140 KB

Ars Pro Vita - the br…
5,472 × 3,648; 1.85 MB

Avent-mayna-trean…
150 × 214; 7 KB

Boloengo.jpg
900 × 1,600; 307 KB

Casco comm.jpg
400 × 300; 15 KB

Casco commemora…
800 × 600; 47 KB

Celestial Buddies.jpg
706 × 807; 160 KB

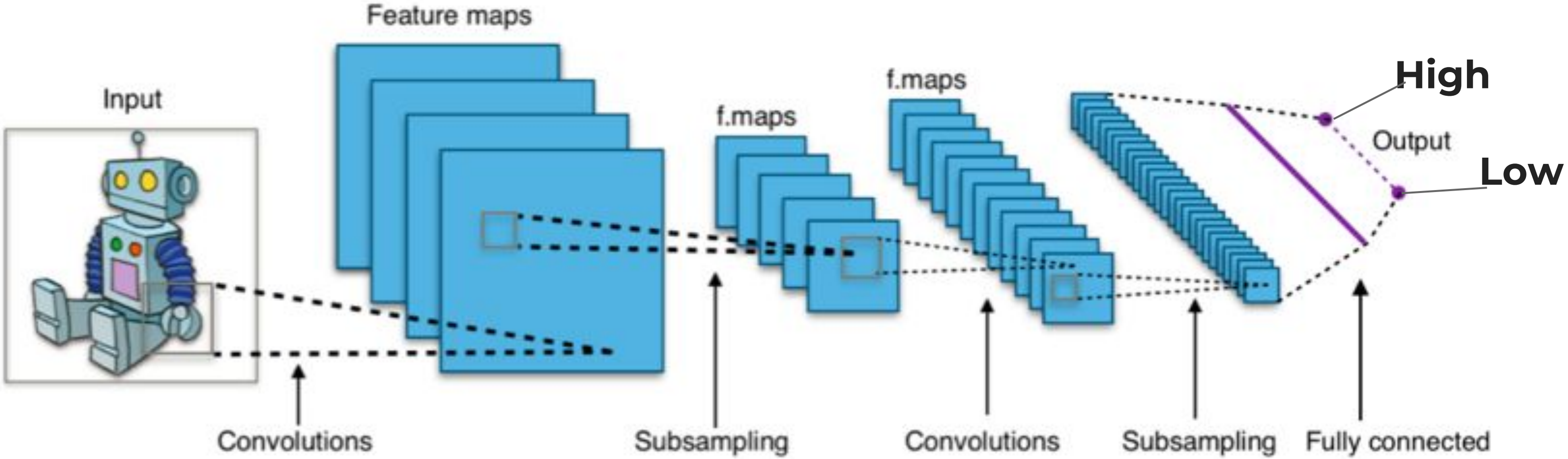
Cibo di sopravviven…
400 × 300; 20 KB

CLONING.jpg
1,472 × 1,536; 565 KB

Lower Quality: 160K
Random Commons

2. **Ranking images** (quality)

FRAMEWORK: Convolutional Neural Network Google Inception-v3[1]

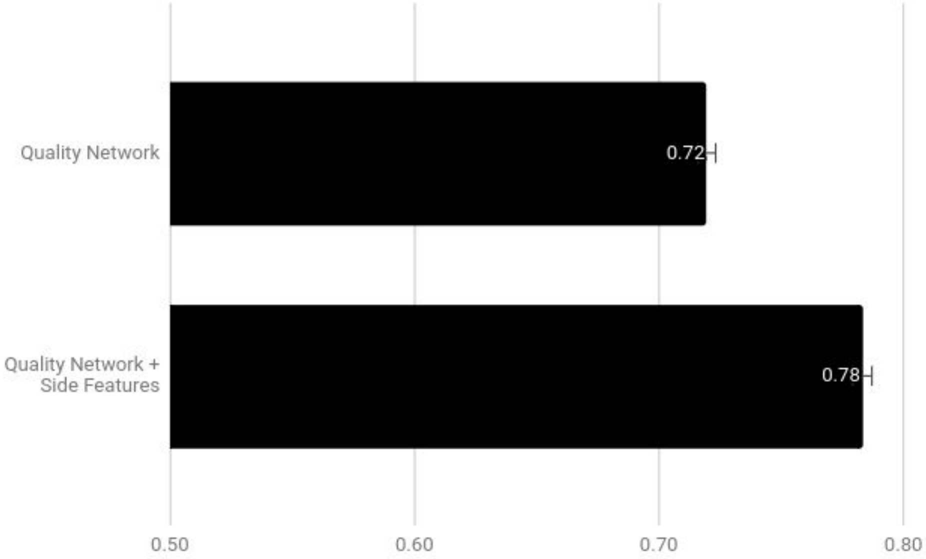


Drawing: Aphex34 on Commons

[1] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

2. **Ranking images** (quality)

RESULTS: ACCURACY on 5-Fold Cross Validation

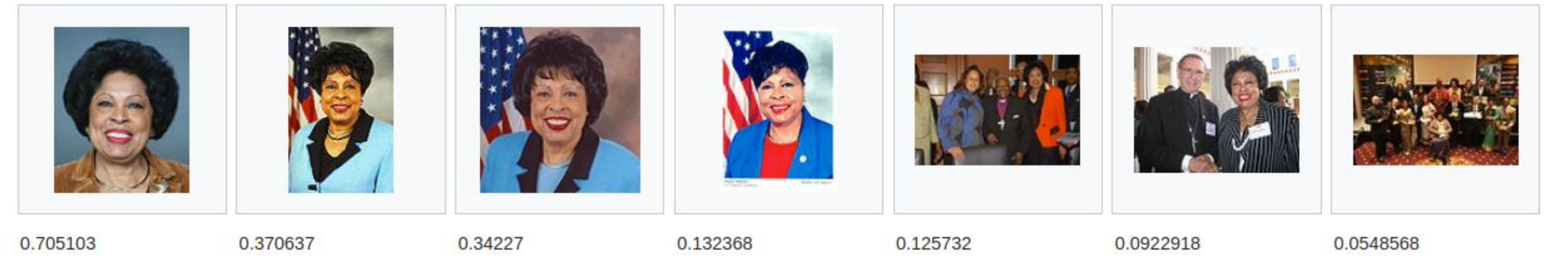


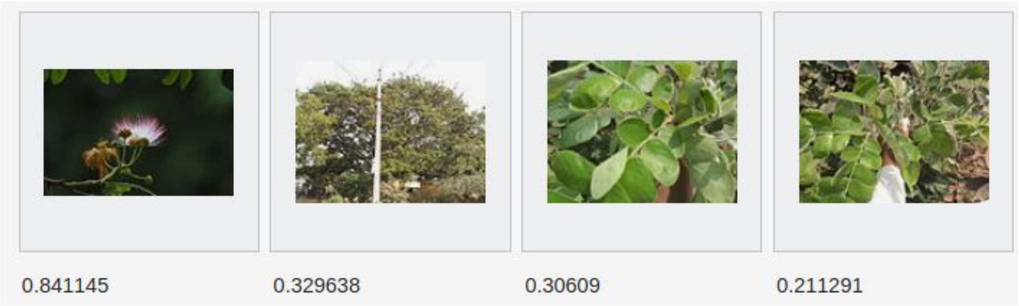
Adding side features:

- Image size
- Description Length

3. **Combining Relevance and Quality**

- **UNSUPERVISED:** Quality based re-ranking of the top X relevant images
- **SUPERVISED:** Joint learning to rank relevance and quality values

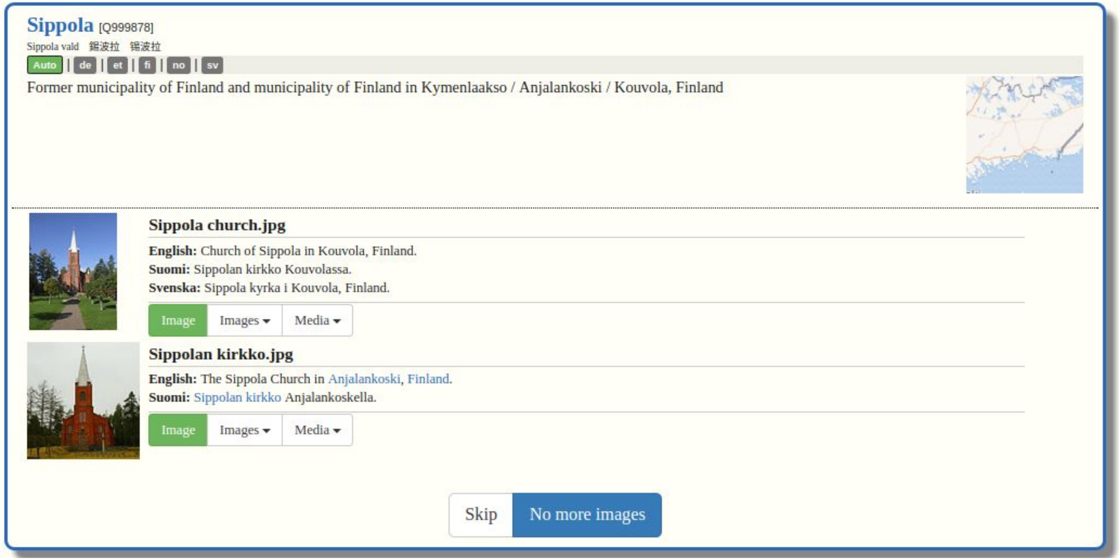




Pantone 7518
5 pt
7.5 pt
10 pt
Stroke 4
17.5 pt
20 pt
11 pt
13 pt
15 pt
Transparent 32
100 g/m²
Transparent 32
100 g/m²
Transparent 50
Transparent 90
Transparent 32
Transparent 31
Transparent 35
PANTONE DE 65-1 C
PANTONE DE 65-2 C
PANTONE DE 65-3 C
PANTONE DE 281-1 C

Evaluation

Offline Evaluation:
The Distributed Game



Sippola [Q999878]

Sippola vald 錫波拉 锡波拉

Former municipality of Finland and municipality of Finland in Kymenlaakso / Anjalankoski / Kouvola, Finland

Sippola church.jpg

English: Church of Sippola in Kouvola, Finland.
Suomi: Sippolan kirkko Kouvolassa.
Svenska: Sippola kyrka i Kouvola, Finland.

Sippolan kirkko.jpg

English: The Sippola Church in Anjalankoski, Finland.
Suomi: Sippolan kirkko Anjalankoskella.

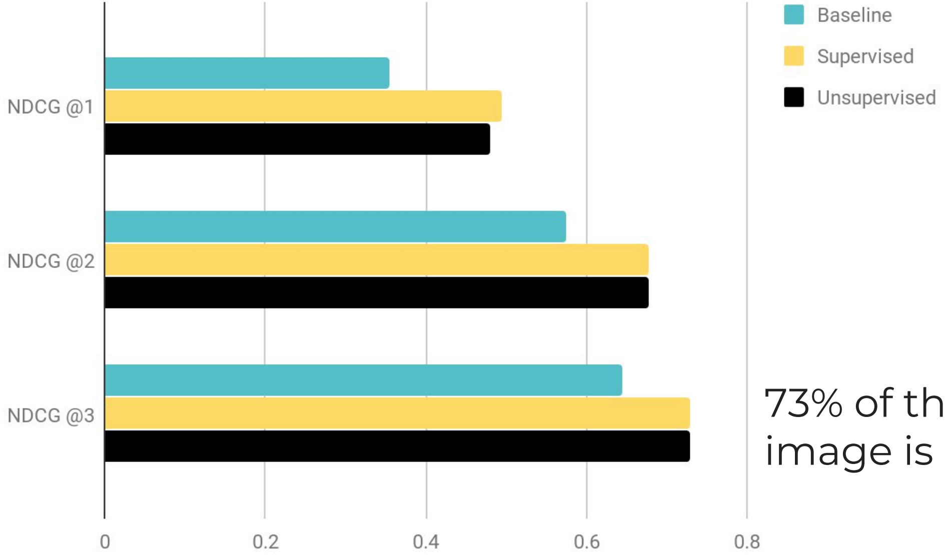
DATA FOR EVALUATION

- 66 K ITEMS with 1+ image candidates
- **User selection** of the item image
- **Classes:**
  'person' : Q5 ,
  'taxon' : Q16521 ,
  'church' : Q16970 ,
  'railway station' : Q55488 ,
  'mountain' : Q8502 ,
  'building' : Q41176

https://tools.wmflabs.org/wikidata-game/distributed/#game=10

Ranking Evaluation

How good are the models proposed at ranking the selected image in the top-X of the candidate images?

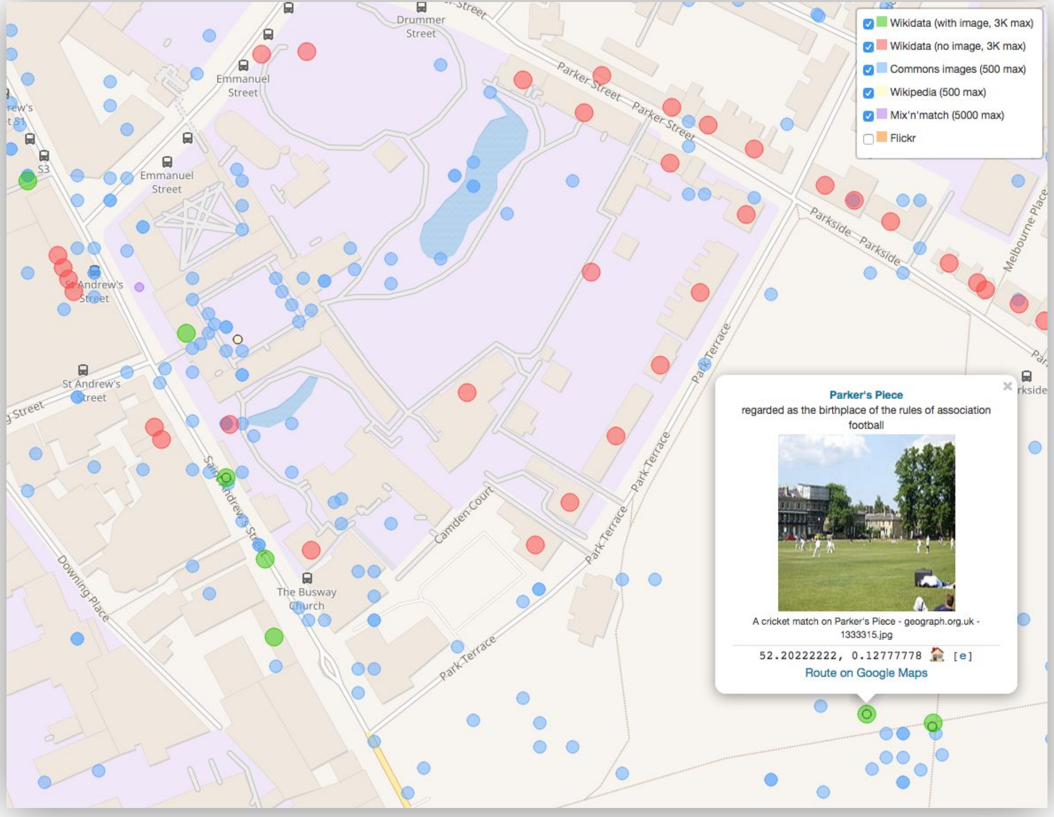


73% of the times the selected image is in the top 3!

Tools for Enriching Wikidata

Magnus Manske

WikiShootMe



☑ 🟩 Wikidata (with image, 3K max)
☑ 🟥 Wikidata (no image, 3K max)
☑ 🟦 Commons images (500 max)
☑ 🟨 Wikipedia (500 max)
☑ 🟪 Mix'n'match (5000 max)
□ 🟧 Flickr

Parker's Piece
regarded as the birthplace of the rules of association football

A cricket match on Parker's Piece - geograph.org.uk - 1333315.jpg

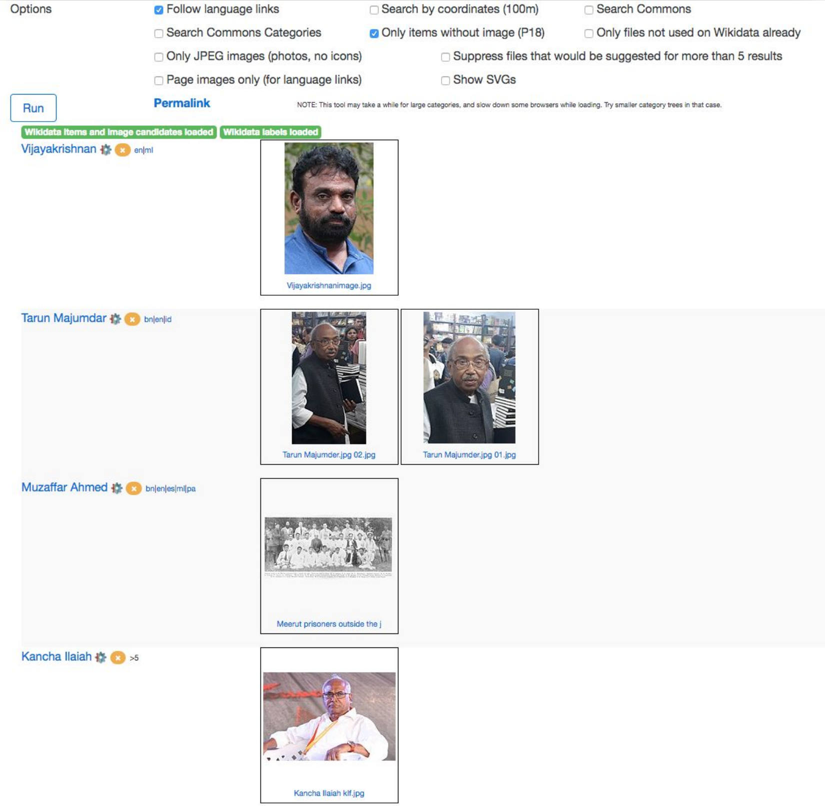
52.20222222, 0.12777778 [e]

Route on Google Maps

- **Map-based tool**
- Shows Wikidata, Wikipedia, Commons images, and other sources on a single map
- **Allows direct upload/adding of images to Wikidata**
- Mobile friendly

https://tools.wmflabs.org/wikishootme/

WD-FIST (WikiData Free Image Search Tool)



Options
☑ Follow language links
□ Search by coordinates (100m)
□ Search Commons
□ Search Commons Categories
☑ Only items without image (P18)
□ Only files not used on Wikidata already
□ Only JPEG images (photos, no icons)
□ Suppress files that would be suggested for more than 5 results
□ Page images only (for language links)
□ Show SVGs

NOTE: This tool may take a while for large categories, and slow down some browsers while loading. Try smaller category trees in that case.

Wikidata items and image candidates loaded
Wikidata labels loaded

Vijayakrishnan en|ml
Vijayakrishnanimage.jpg

Tarun Majumdar bn|en|id
Tarun Majumder.jpg 02.jpg
Tarun Majumder.jpg 01.jpg

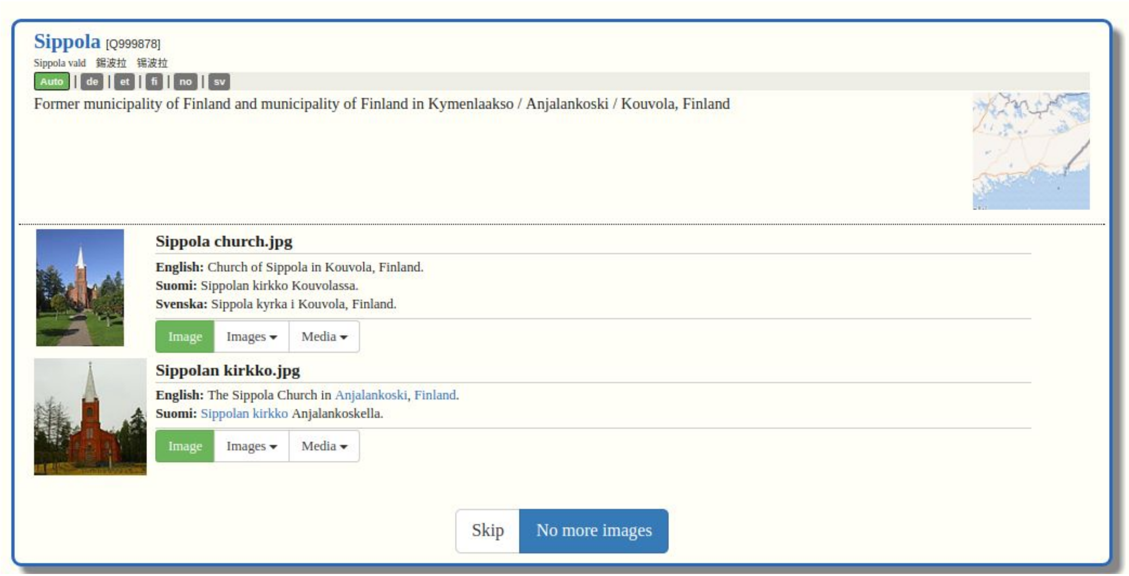
Muzaffar Ahmed bn|en|es|ml|pa
Meerut prisoners outside the j

Kancha Ilaiah >5
Kancha Ilaiah klf.jpg

- Checks items without images
- SPARQL, Categories, lists, etc.
- Uses associated Wikipedia pages, Commons free-text search, GPS etc. to find candidate files
- One-click interface to set an image, two clicks for other media properties
- Can be pre-filled from other tools (e.g. WikiShootMe) or Wikipedia (e.g. Listeria lists)

https://tools.wmflabs.org/fist/wdfist

The Distributed Game



Sippola [Q999878]

Sippola vald 錫波拉 锡波拉

Former municipality of Finland and municipality of Finland in Kymenlaakso / Anjalankoski / Kouvola, Finland

Sippola church.jpg

English: Church of Sippola in Kouvola, Finland.
Suomi: Sippolan kirkko Kouvolassa.
Svenska: Sippola kyrka i Kouvola, Finland.

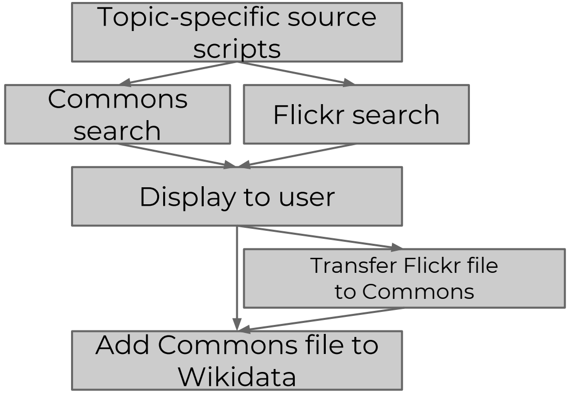
Sippolan kirkko.jpg

English: The Sippola Church in Anjalankoski, Finland.
Suomi: Sippolan kirkko Anjalankoskella.

https://tools.wmflabs.org/wikidata-game/distributed/#game=10

Ongoing Online Evaluation: *File Candidates*

- Pre-computed file-to-item candidates
- “Expensive” queries (e.g. full-text searches)
- Multiple sources (Commons, Flickr, ...)
- Quick, unified display to user
- Specific topics
- Single-click transfer to Commons
- Single click to set as item image



Topic-specific source scripts

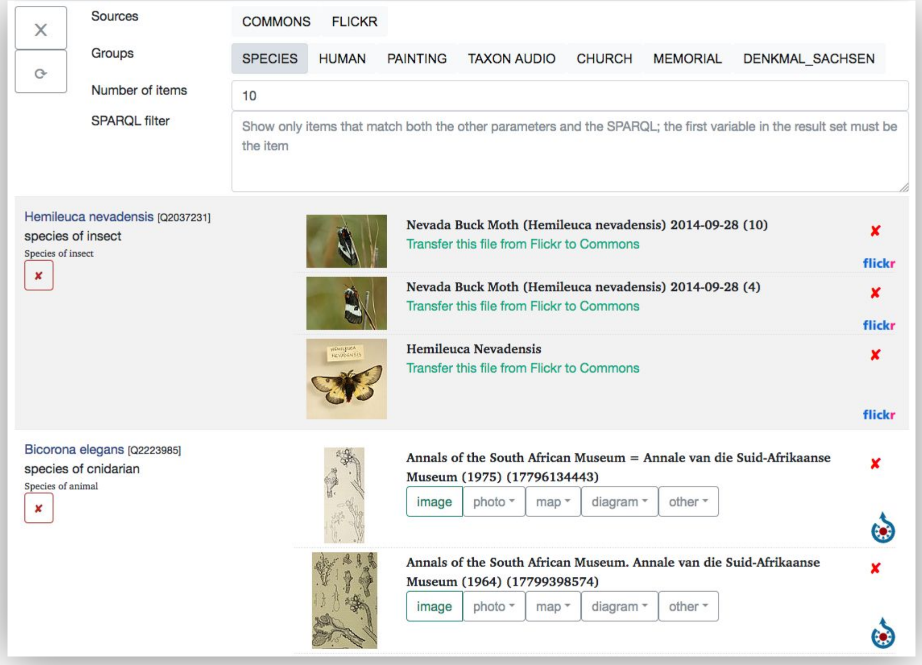
Commons search

Flickr search

Display to user

Transfer Flickr file to Commons

Add Commons file to Wikidata



Sources: COMMONS FLICKR

Groups: SPECIES HUMAN PAINTING TAXON AUDIO CHURCH MEMORIAL DENKMAL_SACHSEN

Number of items: 10

SPARQL filter: Show only items that match both the other parameters and the SPARQL; the first variable in the result set must be the item

Hemileuca nevadensis [Q2037231]
species of insect
Species of insect

Nevada Buck Moth (Hemileuca nevadensis) 2014-09-28 (10)
Transfer this file from Flickr to Commons

Nevada Buck Moth (Hemileuca nevadensis) 2014-09-28 (4)
Transfer this file from Flickr to Commons

Hemileuca Nevadensis
Transfer this file from Flickr to Commons

Bicorona elegans [Q2223985]
species of cnidarian
Species of animal

Annals of the South African Museum = Annale van die Suid-Afrikaanse Museum (1975) (17796134443)

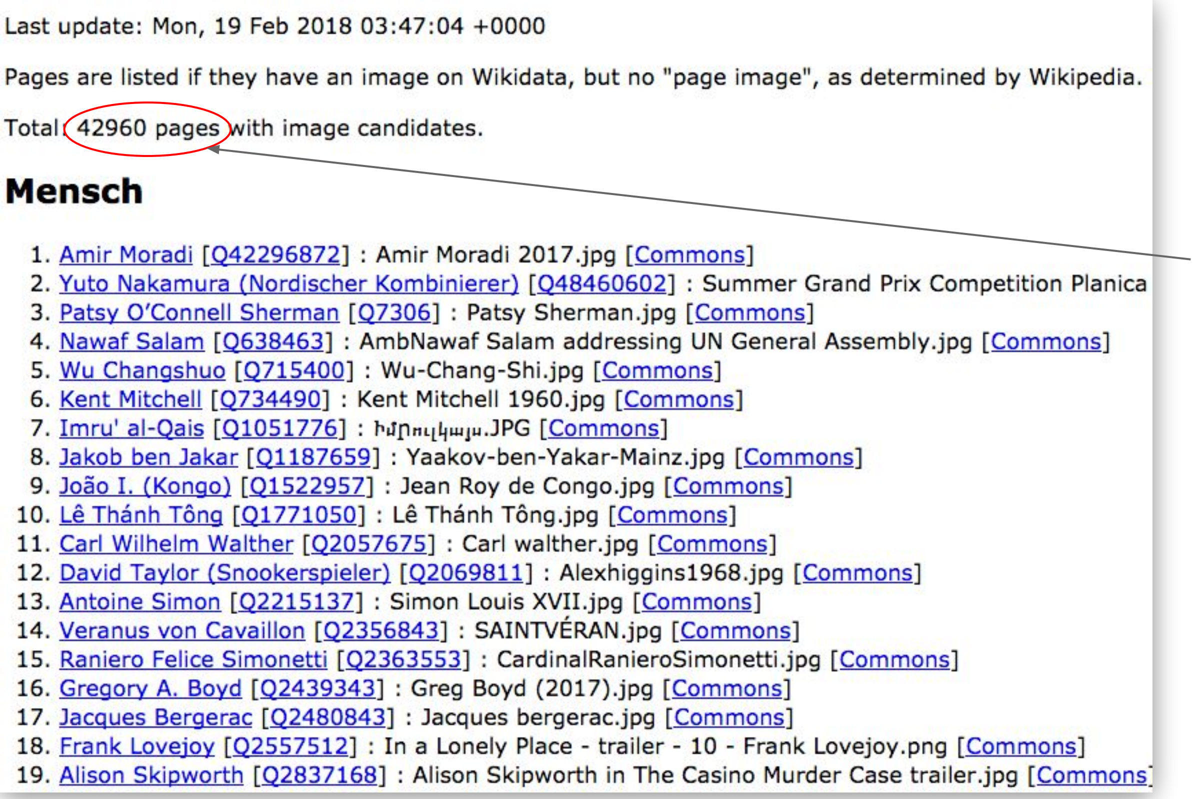
Annals of the South African Museum. Annale van die Suid-Afrikaanse Museum (1964) (17799398574)

https://tools.wmflabs.org/fist/file_candidates/

TO SOMEONE SPECIAL
always smile!
PONY BROWN
HELIOS-44M-4
minolta
XG9

Next!

Back-port images to Wikipedia



Last update: Mon, 19 Feb 2018 03:47:04 +0000

Pages are listed if they have an image on Wikidata, but no "page image", as determined by Wikipedia.

Total: 42960 pages with image candidates.

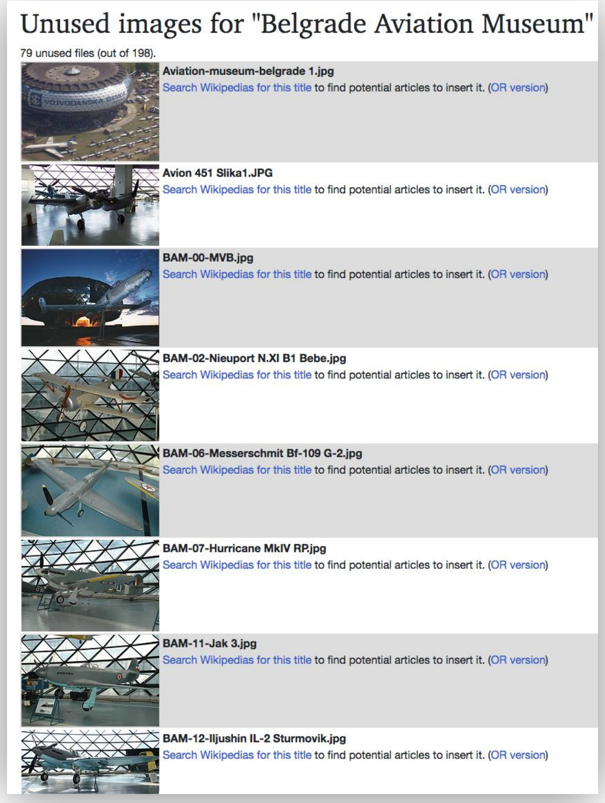
Mensch

1. Amir Moradi [Q42296872] : Amir Moradi 2017.jpg [Commons]
2. Yuto Nakamura (Nordischer Kombinierer) [Q48460602] : Summer Grand Prix Competition Planica
3. Patsy O'Connell Sherman [Q7306] : Patsy Sherman.jpg [Commons]
4. Nawaf Salam [Q638463] : AmbNawaf Salam addressing UN General Assembly.jpg [Commons]
5. Wu Changshuo [Q715400] : Wu-Chang-Shi.jpg [Commons]
6. Kent Mitchell [Q734490] : Kent Mitchell 1960.jpg [Commons]
7. Imru' al-Qais [Q1051776] : Իմրուլկայս.JPG [Commons]
8. Jakob ben Jakar [Q1187659] : Yaakov-ben-Yakar-Mainz.jpg [Commons]
9. João I. (Kongo) [Q1522957] : Jean Roy de Congo.jpg [Commons]
10. Lê Thánh Tông [Q1771050] : Lê Thánh Tông.jpg [Commons]
11. Carl Wilhelm Walther [Q2057675] : Carl walther.jpg [Commons]
12. David Taylor (Snookerspieler) [Q2069811] : Alexhiggins1968.jpg [Commons]
13. Antoine Simon [Q2215137] : Simon Louis XVII.jpg [Commons]
14. Veranus von Cavaillon [Q2356843] : SAINTVÉRAN.jpg [Commons]
15. Raniero Felice Simonetti [Q2363553] : CardinalRanieroSimonetti.jpg [Commons]
16. Gregory A. Boyd [Q2439343] : Greg Boyd (2017).jpg [Commons]
17. Jacques Bergerac [Q2480843] : Jacques bergerac.jpg [Commons]
18. Frank Lovejoy [Q2557512] : In a Lonely Place - trailer - 10 - Frank Lovejoy.png [Commons]
19. Alison Skipworth [Q2837168] : Alison Skipworth in The Casino Murder Case trailer.jpg [Commons]

Pages on German Wikipedia where Wikidata has an image but Wikipedia does not

https://tools.wmflabs.org/wikidata-todo/wp_no_image/dewiki.html

Using unused files



Unused images for "Belgrade Aviation Museum"

79 unused files (out of 198).

Aviation-museum-belgrade 1.jpg
Search Wikipedias for this title to find potential articles to insert it. (OR version)

Avion 451 Slika1.JPG
Search Wikipedias for this title to find potential articles to insert it. (OR version)

BAM-00-MVB.jpg
Search Wikipedias for this title to find potential articles to insert it. (OR version)

BAM-02-Nieuport N.XI B1 Bebe.jpg
Search Wikipedias for this title to find potential articles to insert it. (OR version)

BAM-06-Messerschmit Bf-109 G-2.jpg
Search Wikipedias for this title to find potential articles to insert it. (OR version)

BAM-07-Hurricane MkIV RP.jpg
Search Wikipedias for this title to find potential articles to insert it. (OR version)

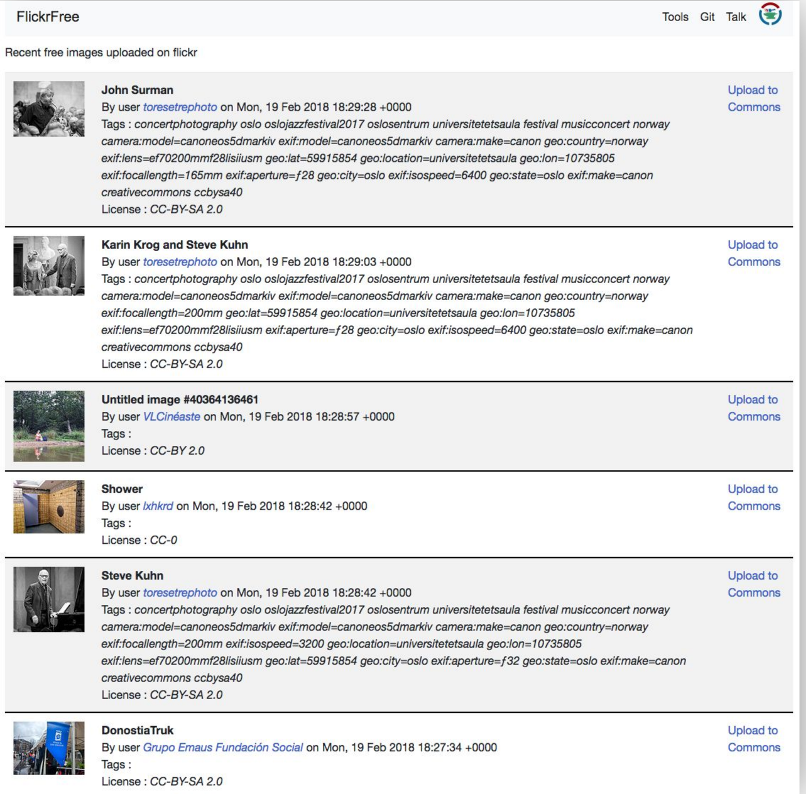
BAM-11-Jak 3.jpg
Search Wikipedias for this title to find potential articles to insert it. (OR version)

BAM-12-Iljushin IL-2 Sturmovik.jpg
Search Wikipedias for this title to find potential articles to insert it. (OR version)

- Many great images already on Commons, some by donation
- Takes a category on Commons, and shows all files *not* used on any Wikipedia/Wikidata/etc.
- Might even **spark ideas for new articles or items**

https://tools.wmflabs.org/magnustools/unused_files.php

Flickr firehose



Recent free images uploaded on flickr

John Surman
By user *toresetrephoto* on Mon, 19 Feb 2018 18:29:28 +0000
Tags : *concertphotography oslo oslojazzfestival2017 oslosentrum universitetetsaula festival musicconcert norway camera:model=canoneos5dmarkiv exif:model=canoneos5dmarkiv camera:make=canon geo:country=norway exif:lens=ef70200mmf28lisiiusm geo:lat=59915854 geo:location=universitetetsaula geo:lon=10735805 exif:focallength=165mm exif:aperture=ƒ28 geo:city=oslo exif:isospeed=6400 geo:state=oslo exif:make=canon creativecommons ccbysa40*
License : *CC-BY-SA 2.0*
Upload to Commons

Karin Krog and Steve Kuhn
By user *toresetrephoto* on Mon, 19 Feb 2018 18:29:03 +0000
Tags : *concertphotography oslo oslojazzfestival2017 oslosentrum universitetetsaula festival musicconcert norway camera:model=canoneos5dmarkiv exif:model=canoneos5dmarkiv camera:make=canon geo:country=norway exif:focallength=200mm geo:lat=59915854 geo:location=universitetetsaula geo:lon=10735805 exif:lens=ef70200mmf28lisiiusm exif:aperture=ƒ28 geo:city=oslo exif:isospeed=6400 geo:state=oslo exif:make=canon creativecommons ccbysa40*
License : *CC-BY-SA 2.0*
Upload to Commons

Untitled image #40364136461
By user *VLCinéaste* on Mon, 19 Feb 2018 18:28:57 +0000
Tags :
License : *CC-BY 2.0*
Upload to Commons

Shower
By user *lxhkrd* on Mon, 19 Feb 2018 18:28:42 +0000
Tags :
License : *CC-0*
Upload to Commons

Steve Kuhn
By user *toresetrephoto* on Mon, 19 Feb 2018 18:28:42 +0000
Tags : *concertphotography oslo oslojazzfestival2017 oslosentrum universitetetsaula festival musicconcert norway camera:model=canoneos5dmarkiv exif:model=canoneos5dmarkiv camera:make=canon geo:country=norway exif:focallength=200mm exif:isospeed=3200 geo:location=universitetetsaula geo:lon=10735805 exif:lens=ef70200mmf28lisiiusm geo:lat=59915854 geo:city=oslo exif:aperture=ƒ32 geo:state=oslo exif:make=canon creativecommons ccbysa40*
License : *CC-BY-SA 2.0*
Upload to Commons

DonostiaTruk
By user *Grupo Emaus Fundación Social* on Mon, 19 Feb 2018 18:27:34 +0000
Tags :
License : *CC-BY-SA 2.0*
Upload to Commons

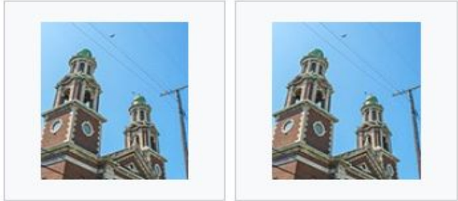
- Flickr version of “Recent Changes”
- Shows latest images on Flickr with a free license
- Upload button to Commons
- Gets us new images for the other tools

https://tools.wmflabs.org/magnustools/flickrfree.php

Finding Free-Licensed images

- Are **Flickr images already on Commons?**
- **Computer vision** can detect **exact and near-duplicate images.**
- In a small scale experiment, we found that only **0,1% of these images are already on Commons.**

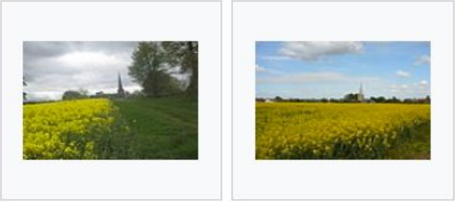
Exact Duplicates



Commons

Flickr

Near-Duplicates



Commons

Flickr

Future Paths

- Improve machine learning models for **image recommendation** (help reducing the backlog!)
- Leverage efforts on Wikidata to **make Wikipedia more visual**
- Acquire more free-licensed visual data from **external sources**

Inspire readers and editors through images

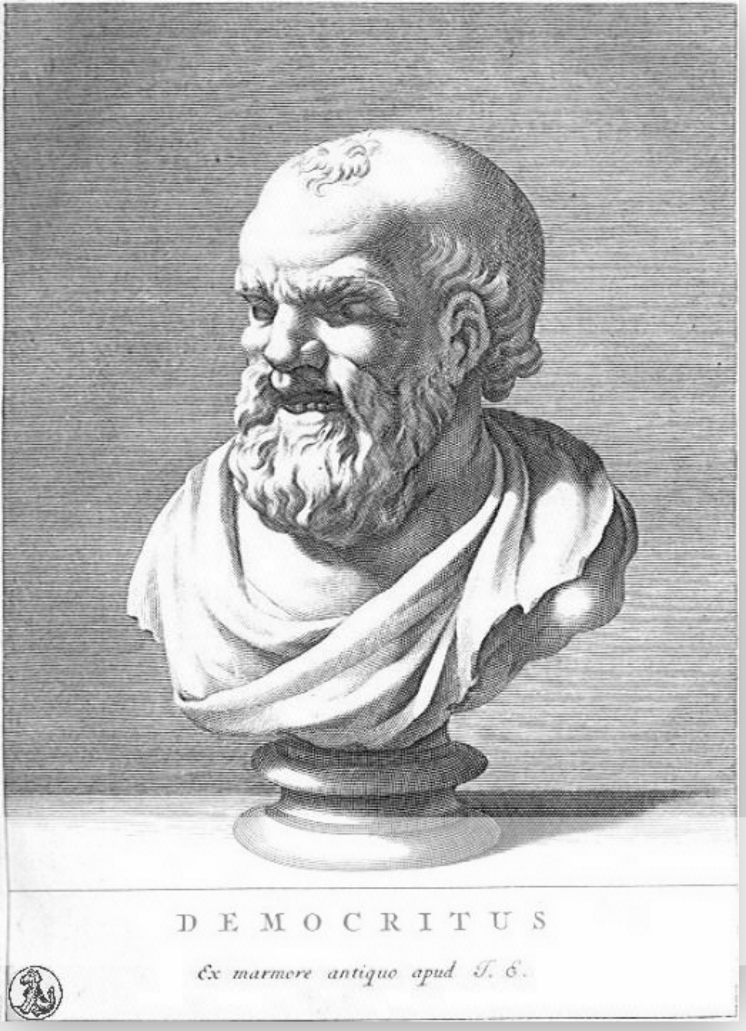
Thank you :)

miriam@wikimedia.org

magnusmanske@googlemail.com

https://meta.wikimedia.org/wiki/Research:Recommending_Images_to_Wikidata_Items

WIKIMEDIA
FOUNDATION



Visual **Thinking**