Explore large datasets of images with

Panoptic

Panoptic **A**chieves **N**umerical **O**rganisation of **P**ictures **T**hrough **I**nteractive **C**omputing

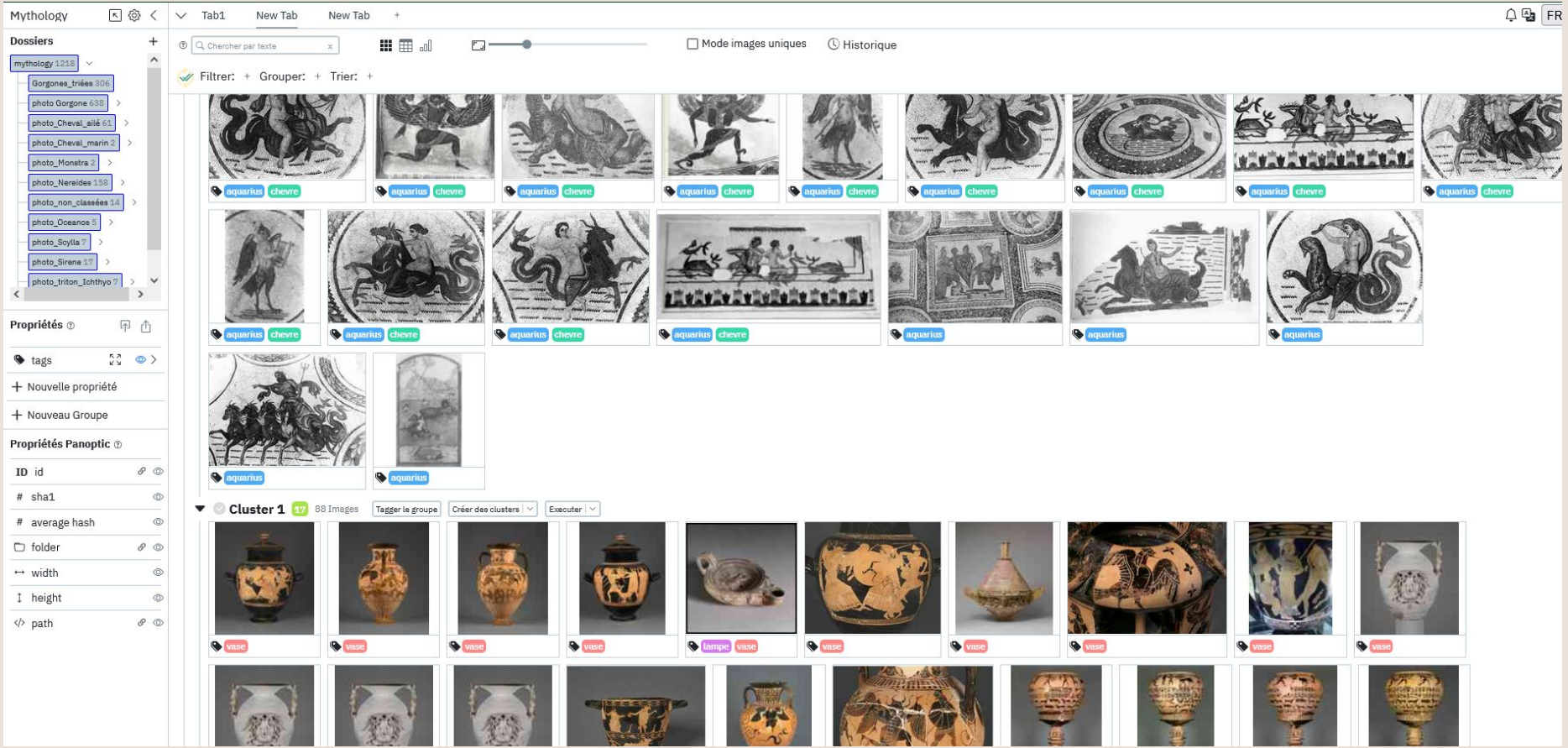
FOSDEM - Open Research
01/02/2025

Félix Alié (CERES, Sorbonne Université)
Édouard Bouté (CERES, Sorbonne Université)
David Gödicke (CERES, Sorbonne Université)

Summary

- Introduction
- What’s panoptic
- Research Use Cases
- Architecture and Plugins

What’s Panoptic ?


Mythology
Tab1
New Tab
New Tab
FR
Dossiers
mythology 1218
Gorgones_triées 306
photo Gorgone 638
photo_Cheval_ailé 61
photo_Cheval_marin 2
photo_Monstra 2
photo_Nereides 158
photo_non_classées 14
photo_Oceanos 5
photo_Scylla 7
photo_Sirene 17
photo_triton_Ichthyo 7
Chercher par texte
Mode images uniques
Historique
Filtrer:
Grouper:
Trier:
Propriétés
tags
Nouvelle propriété
Nouveau Groupe
Propriétés Panoptic
ID id
sha1
average hash
folder
width
height
path
aquarius
chevre
Cluster 1
17
88 Images
Tagger le groupe
Créer des clusters
Executer
vase
lampe

A bit of context

- Software for dynamic exploration of medium and large datasets : 100 → 100k (1M?) images on YOUR computer.
- In development since May 2023.
- Open source: everything is on Github.
- Two software engineers, a research designer, and a humanities research engineer

Origin

- Media studies researchers working on Twitter image datasets too big for manual analysis
- How to analyze the circulation of similar images, even when they are slightly differents ?
- Hard to know exactly what’s in a big dataset.
- Tropy is great for manual analysis but lack of automatic assistance
- Great ML tools existing but need to know how to code and not easily usable to explore and annotate along

➔ Need of a real exploration UI assisted with ML tools

Quick Demo

- Properties
- Filters, Groups, Sorts
- Clustering
- Similarity
- Recommandations
- Export everything

some Panoptic use cases

Current use cases

- Exploration of large web corpora
- Exploration of large digitized corpora
- Exploration of film corpora

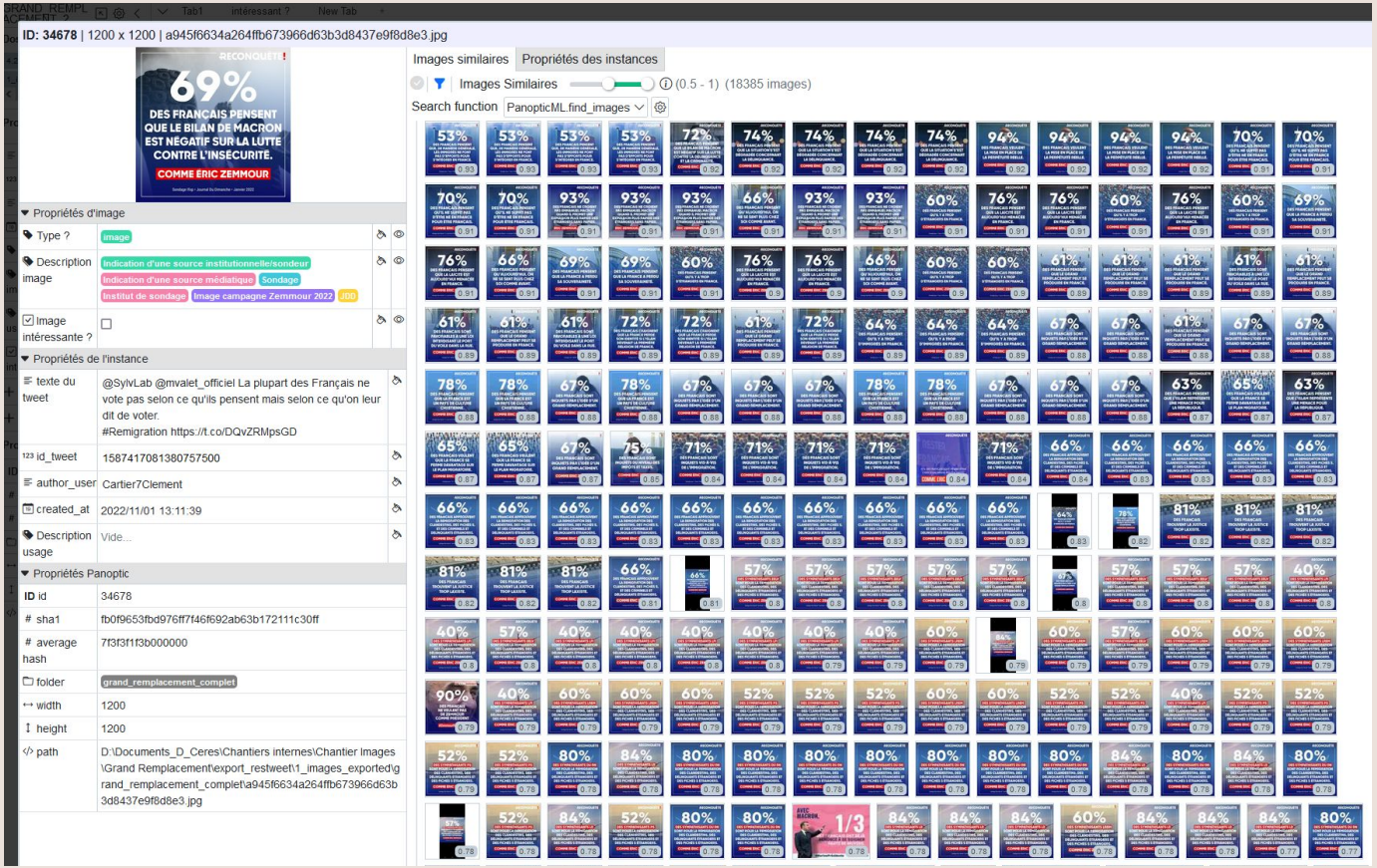
Exploration of large web corpora

Examples :

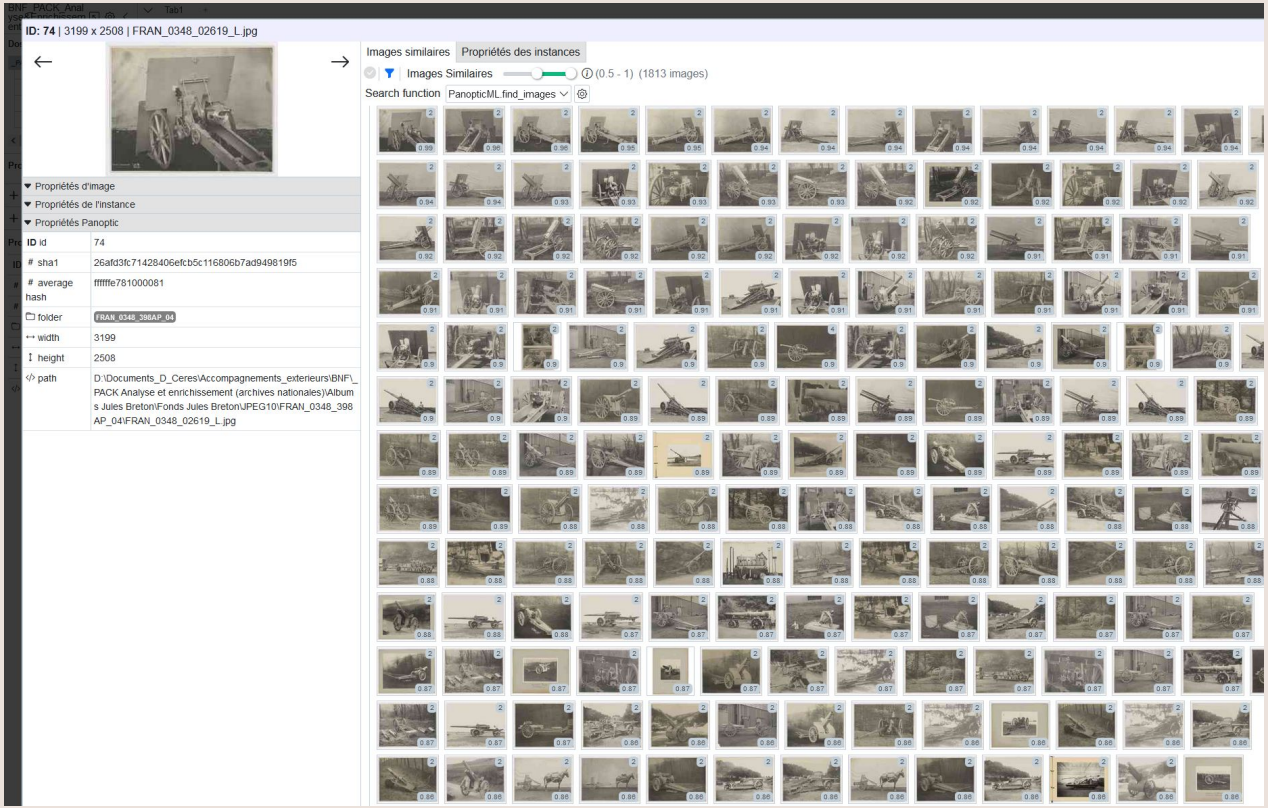
Social network	*Subject*	*Number of single images*	*Number of publications*
Twitter & Facebook	The circulation and problematization of images of police violence on social networks during the Yellow Vests protests.	31 698	301 655
Twitter	The spread of the racist ideology of “grand remplacement” on social networks and its vernacular visuals	50 783	145 094
Twitter	Development of specific vernacular visuals by the anti-genre movement between 2013 and 2021	19 703	139 815
Instagram	The fan culture around Eurovision	64 702	65 814
TikTok	Information production formats for news stories and their politicization	6 234	6 234

Exploration of large web corpora

Edouard Bouté's thesis
& CERES corpus

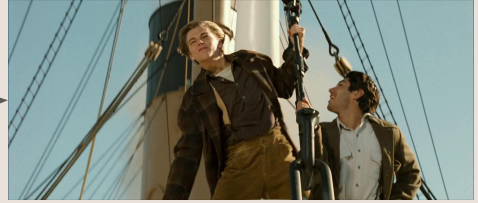


Exploration of large digitized corpora



BNF's archives: Fonds Jules Breton

Exploration of film corpora

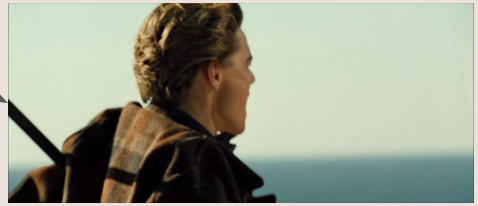


Titanic_00h_32m_19s.jpg

Titanic_00h_32m_24s.jpg

Titanic_00h_32m_29s.jpg

Titanic_00h_32m_34s.jpg



Titanic_00h_32m_39s.jpg

Titanic_00h_32m_44s.jpg

Titanic_00h_32m_49s.jpg

Titanic_00h_32m_54s.jpg

→ Spot and identify the visual codes of a filmic genre, for example, here, repeated shot compositions throughout my corpus

Becoming Jane (2007), *Emma*. (2020
Lost in Austen (2008), *Northanger Abbey* (2007)
Pride and Prejudice (2005), *Pride and Prejudice* (1995)

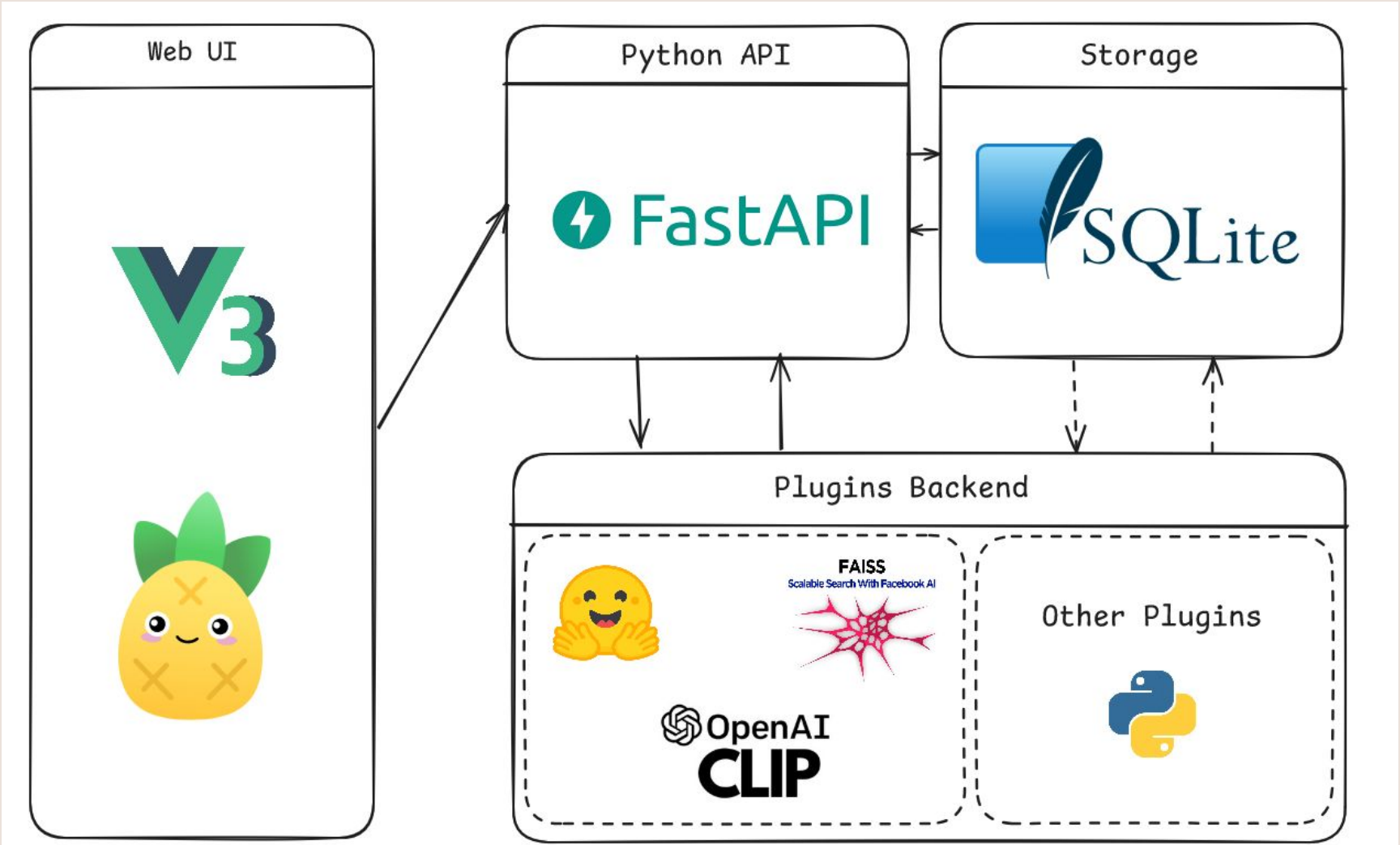
Léa Andolfi's thesis (Sorbonne Université)

Architecture and Plugins

Panoptic Architecture

Requirements:

- Runs on Linux / Windows / MacOs
- Should be “easy” to install: pip install panoptic
- Allow plugins for ML algorithms
- Work locally


Web UI
Python API
FastAPI
Storage
SQLite
Plugins Backend
FAISS
Scalable Search With Facebook AI
Other Plugins
OpenAI
CLIP

Database: Sqlite

- No DB server to setup
- One File contains all data
- Easy to share / backup
- SQL based

Backend: Python

- Multi OS support
- Easy to load plugins
- PyTorch

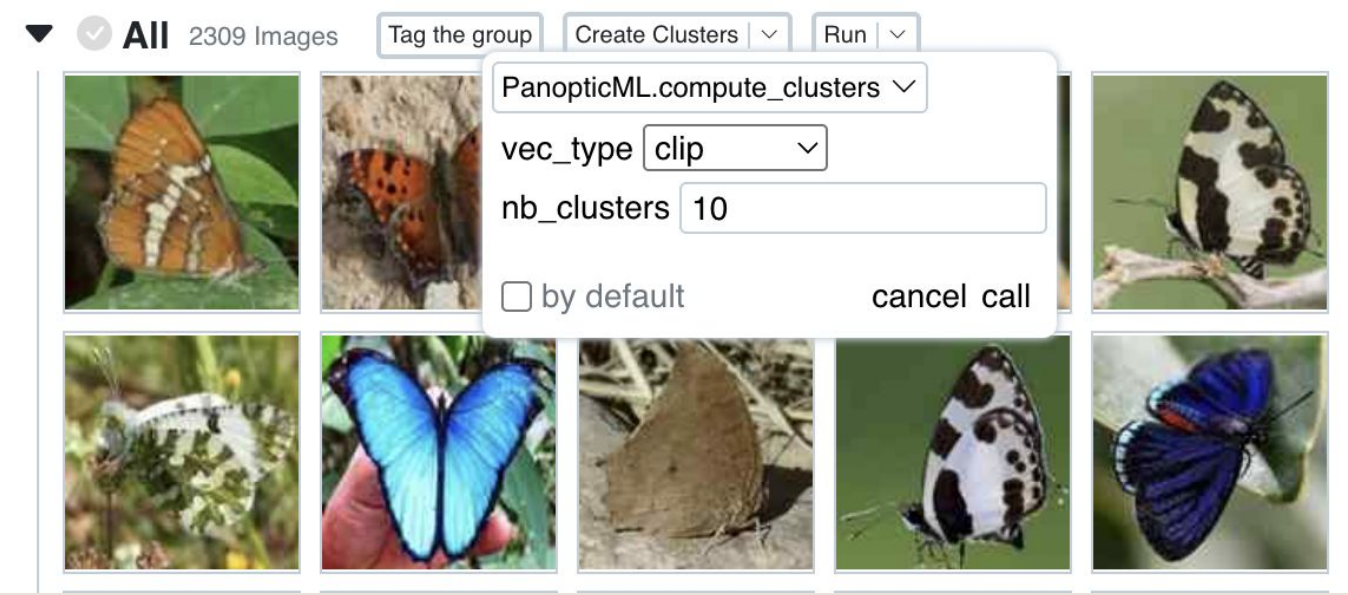
Frontend: WebUI (vue3)

- Everyone has a browser
- Unified UI experience on every OS
- Remote access by default
- No heavy UI library to install

Plugins

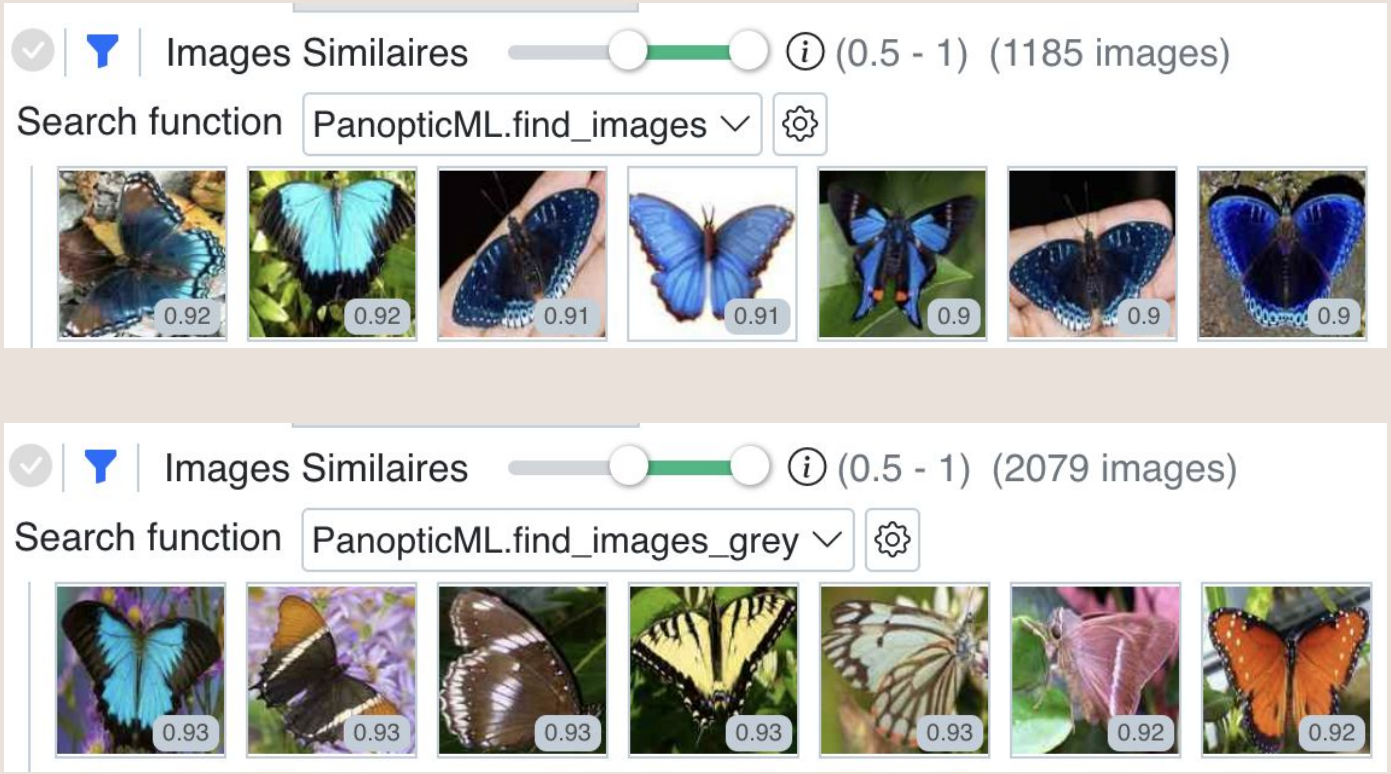
- Customize Panoptic for your needs
- Possible Actions: Cluster, Similarity, Execute
- Example: Clustering of Memes: Image + Text

Plugins: Clustering

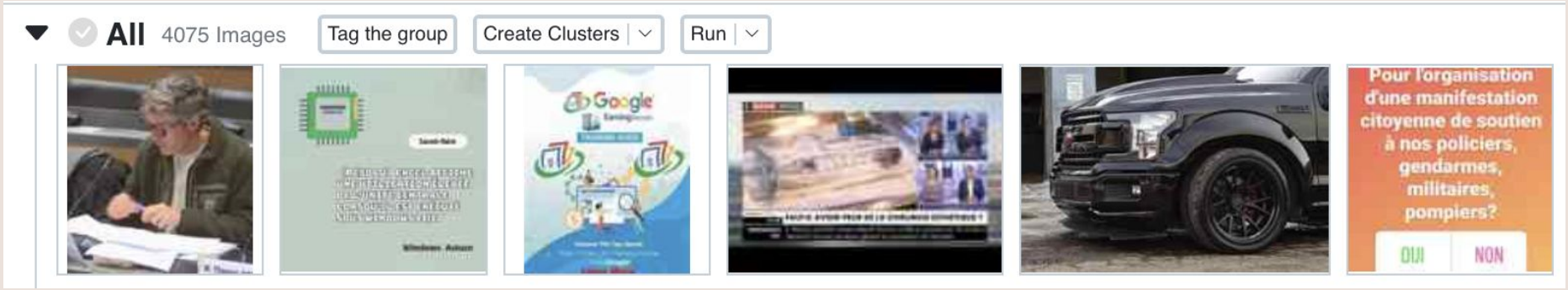



```python
def compute_clusters(self, context: ActionContext, vec_type: VectorType = VectorType.clip, nb_clusters: int = 10):
```

Plugins: Similarity



Plugins: Execute



Questions ?