

# EcoParse

## Automated Species-Level Data Extraction Tool



Authors: Adam Ulicny, Florencia Grattarola, Gabriel Ortega, Ivo Kadlec, Petr Keil  
MOBI lab, Czech University of Life Sciences in Prague, e-mail: ulicny@fld.czu.cz

### The Problem:

Valuable species-level data is often hidden in various sources.

For large-scale analyses, data must be extracted reliably. This can take many hours.

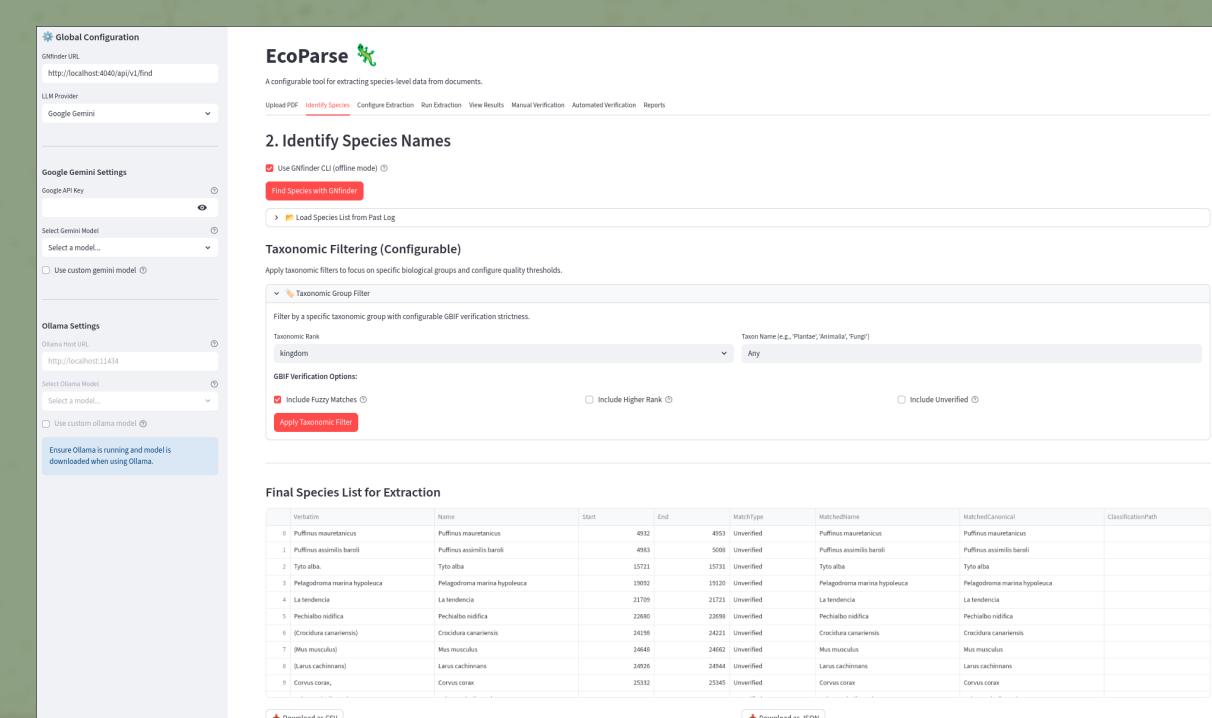
Our case:  
Thousands of Regional Redlists  
**RegRed project**

### Our Solution:

User-friendly browser-based GUI

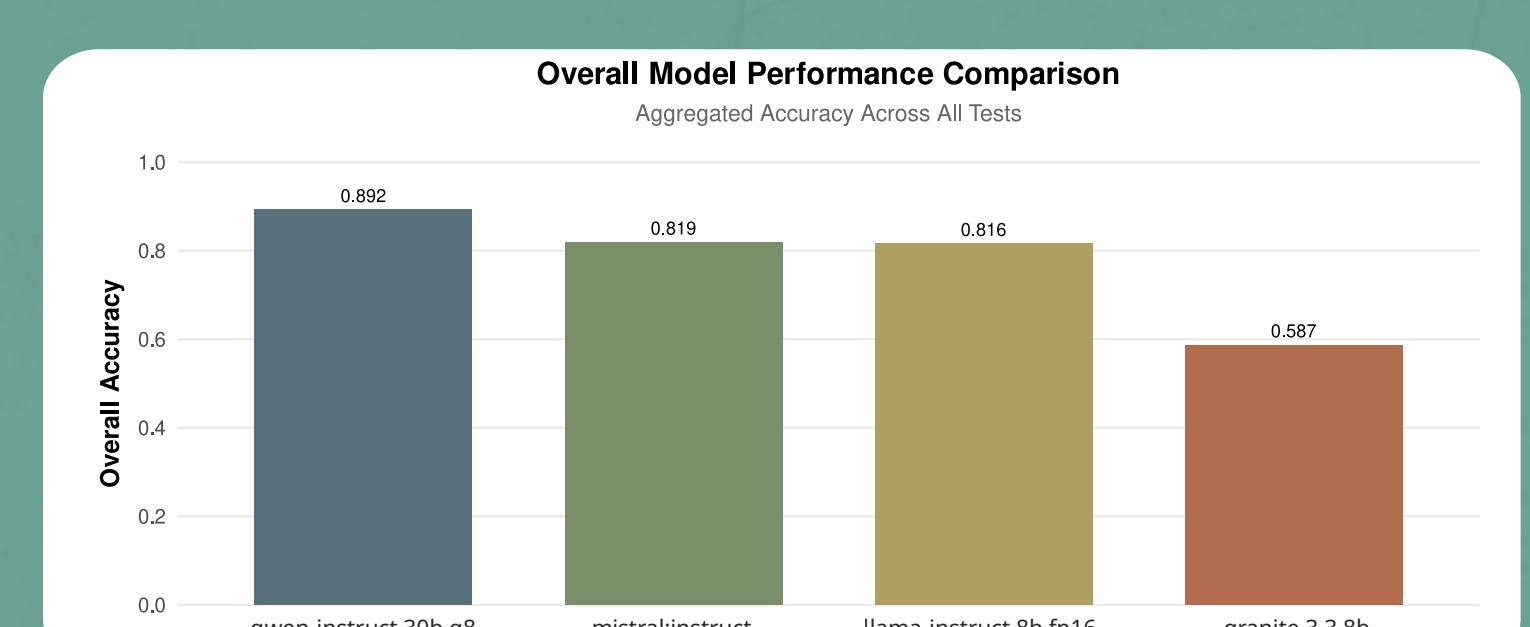
Using **existing LLM's + NER**

Local and Cloud API's for flexibility



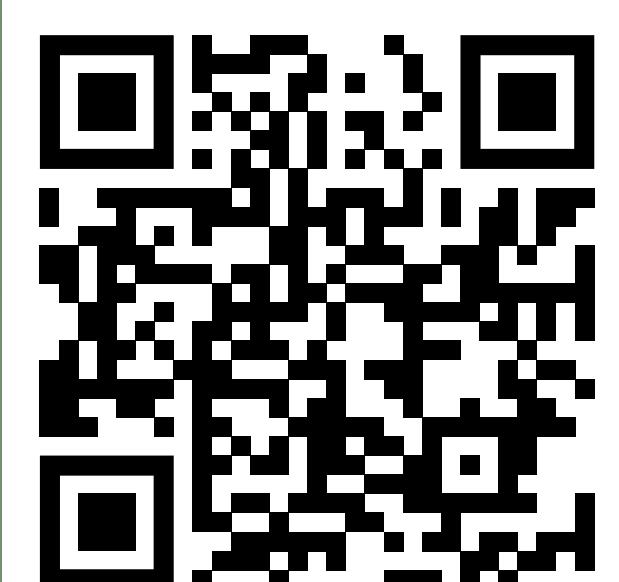
### Some results:

We tested against 8 Redlists  
Compared 4 local models  
Identical settings between runs.  
20GB VRAM GPU



Best performer = **Qwen-instruct**  
~ 89 % mean accuracy

### Learn more:



<https://github.com/AdamUlicny/EcoParse>

### File Loader

Trim unnecessary sections

-Reduce API calls, Token cost & Error rate

Extract and clean text

X Introduction →

Relevant Portion {

X References →



### Species names

#### GNFinder

Locating potential candidates

#### GBIF API

Species verification and filtering

Initial sp name candidates:

Actualmente esta zona  
*Alnus acuminata*  
*Ampelion rufaxilla*  
*Andes orientales*  
*Andigena hypoglauca*  
*Andigena laminirostris*  
*Andigena nigrirostris*  
*Anisognathus notabilis*  
*Antidaphne andina*  
*Ara militaris*

Final filtered species list:

*Ampelion rufaxilla*  
*Andigena hypoglauca*  
*Andigena laminirostris*  
*Andigena nigrirostris*  
*Anisognathus notabilis*  
*Ara militaris*



### Examples

Provide example input/output

Explain where to find data

#### Few-shot prompting

Vuur salamander (*Salamandra salamandra* ssp. *terrestris*)

Rode Lijst 2023: ernstig bedreigd  
Rode Lijst 2007: bedreigd  
IUCN Nederland 2023: critically endangered

Nederlandse criteria

Zeldzaamheid: de soort leeft niet tussen de 50 en 250 voorvalen. De soort plant zich vooral in een habitat waar, wat beide gevallen leidt tot zeldzaamheidsklasse zeer zeldzaam (zz).

Trend: sinds 1950: de verspreiding is afgenomen met 50% wat leidt tot trendklasse sterk afgenomen (tt). De populatiegrootte is afgenomen met 90%, wat leidt tot de zwaardere trendklasse zeer sterk afgenomen (rr).



Source: Netherlands Redlist of Reptiles & Amphibians 2023

Input: "IUCN Nederland 2023: critically endangered"

Output: CR

Explainer: Only IUCN 2023 data

Adding examples and explainers can greatly improve extraction accuracy.

Especially useful in ambiguous or complicated cases.

### Chunking

Contextual chunks

Per species

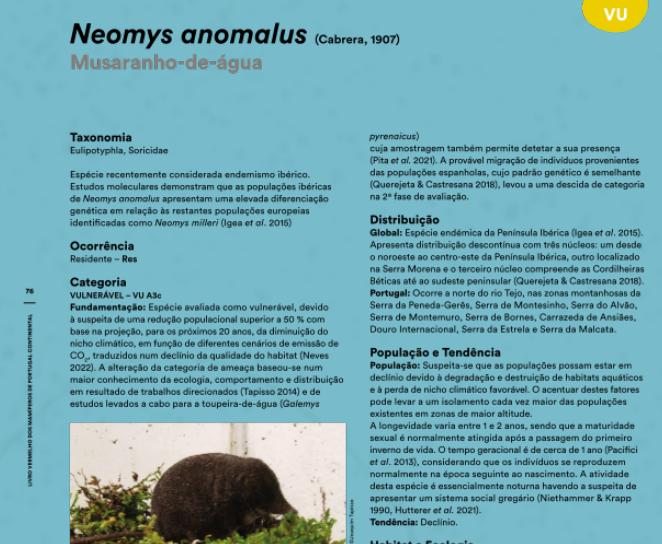
Configurable length

### Text

<Pardela Chica  
Puffinus assimilis baroli  
En Peligro; EN B2ab (i,ii,iii); C2a(ii)  
Autores: Domingo Trujillo y Juan José  
Ramos  
En el archipiélago canario está  
presente la subespecie P. a. baroli,  
habiéndose constatado su>

- speed and cost

### Image



- preserves structure

### LLM parsing

**Ollama** (local) or **Gemini** (cloud)

Pick model according to needs  
**Support for custom models**

Returns **JSON list**

### Verification

Quickly manually verify results

Extracted sections highlighted

### Download Results!

Sample raw output from EcoParse:		
Species	Status	Criteria
<i>Gavia immer</i>	VU	D1
<i>Podiceps nigricollis</i>	NT	WU D1
<i>Podiceps cristatus</i>	NF	NF
<i>Bulweria bulwerii</i>	EN	B2ab(i,ii,iii,iv)
<i>Asio otus</i>	NF	NF
<i>Larus cachinnans</i>	NF	NF
<i>Calonectris diomedea diomedea</i>	EN	A3cde
<i>Bubo bubo</i>	NF	NF
<i>Calonectris diomedea borealis</i>	VU	A3d+4d
<i>Puffinus puffinus</i>	EN	B2ab(i,ii); C2a(ii)
<i>Puffinus mauretanicus</i>	CR	A3ace+4ace; B2ab(i,ii,iv,v); E
<i>Balaenoptera bidens</i>	NF	NF
<i>Puffinus assimilis baroli</i>	EN	B2ab(i,ii); C2a(i)
<i>Pelagodroma marina hypoleuca</i>	VU	NF
<i>Pelecanoides urinatrix</i>	NF	NF

Funded by the European Union (ERC, BEAST, 101044740).

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency.

Neither the European Union nor the granting authority can be held responsible for them.

