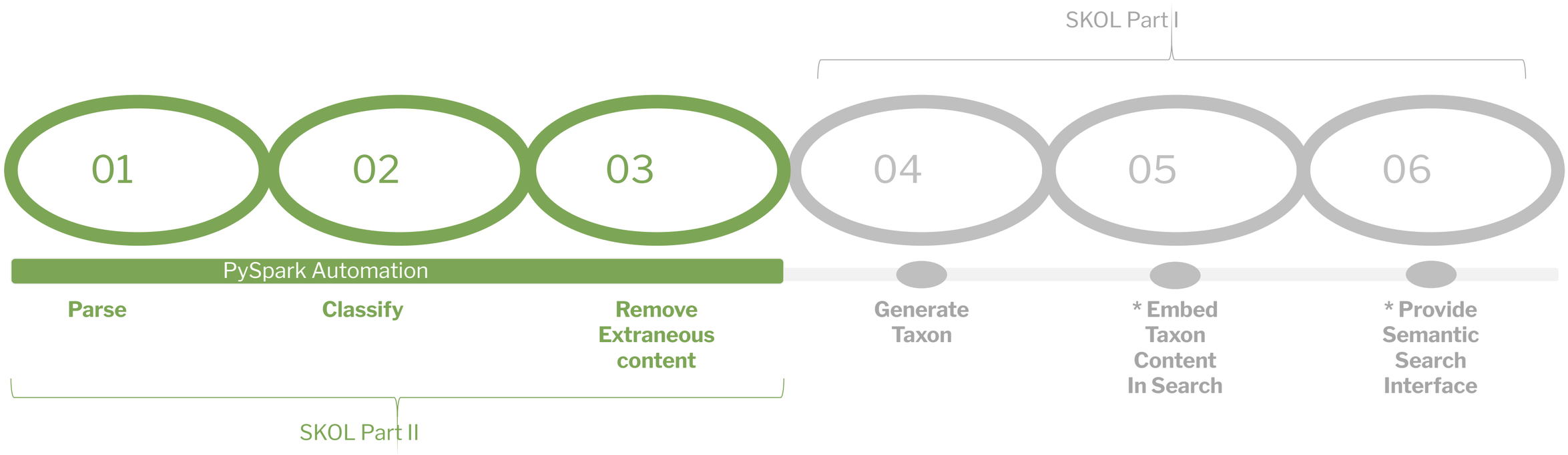
Synoptic Key of Life MASMC 2025
A work in progress



May 2019: Manual annotation

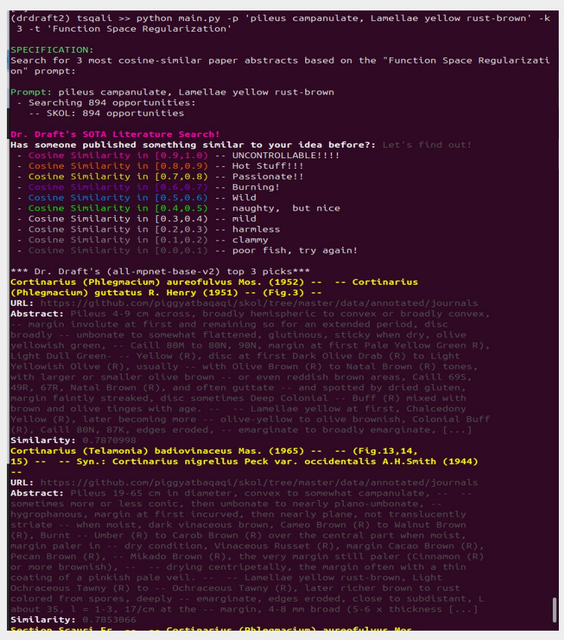
- Annotated 244 issues from old issues of 3 freely available journals:
  - Persoonia, Mycotaxon, and Mycologia
  - from mykoweb.com

```
[@Glomus cubense Y. Rodr. & Dalpé, sp. nov.#Nomenclature*]

[@Plates 1–2#Figure*]

[@Mycobank MB561650#MB*]
[@Sporocarpia ignota. Sporae singulares vel aggregatae. Sporae hyalinae vel luteolae,
ovoideae, vel irregulares, 20–48 × (24–)54–82 µm, raro globosae, 24–54 µm diam.
Sporae tunica stratis duobus; stratum exterius hyalinum, subflexuosum, ad 0.8–1.0 µm
crassum, stratum interius hyalinum vel luteolum, 0.8–1.7 µm crassum, rigidum. Hyphae
sustinentes hyalinae, rectae, subcylindricae vel subinfundibuliformes, 4–10 µm crassae ad
basim sporae. Porus apertus, raro septo curvo clausus cum strato interiore. Mycorrhizas
vesicular-arbusculares formantes.#Description*]
```

Feb 2025 IST664 Language



- Based on Dr. Draft's SOTA LIterature Search from the Auton Lab.
- Uses a Large Language Model (SBERT) to create a space of all nomenclature/description sets.
- Provide a description on the command line which is embedded into the space.
- Returns the closest matching descriptions by cosine similarity.
- Should be publicly available at https://synoptickeyof.life by end of June 2025.

March 2025: IST718: pyspark Classification Pipeline

- over 1000 unannotated documents to ingest
- Parses OCR output into paragraphs
- Classifies paragraphs into Nomenclature, Description, Misc-exposition.
- Distributed pipeline to allow parallel processing.

June 2025: IST691: Feature Extraction for a Synoptic Key

- Uses an LLM (gemma) to extract features and their values from each description
- Annotate the resulting json with TF-IDF scores for every feature
- Basis for building a synoptic key User Interface

Going forward: Productionization

- Web bot to collect open content articles
- Automated OCR pipeline
- Provenance annotation
- Front end

Resources

Collaborators:
JJ. Balasi
D. Caspers
C. Murphy
S. Osuga
L.H.P. Yarroll

2025-09-25

coming soon: https://synoptickeyof.life
https://github.com/piggyatbaqaqi/skol
https://github.com/piggyatbaqaqi/dr-drafts-mycosearch
IST664 Synoptic Key of Life
IST718 Synoptic Key of Life II
https://github.com/autonlab/dr-drafts-sota-literature-search

This poster: