

FORMINDEX_10_13_2024

Daniel Friedman
0000-0001-6232-9096
October 13, 2024

Situation

In [⑧ FORMINDEX_10-08-2024](#) it was identified that the early build of **FORMINDEX scripts** could successfully read in & visualize the FORMIS export Bibtex format. Sanford sent an initial test set of citations, and then the full July 2024 FORMIS export.

Here is the sample of 316 citations

 FORMIS 2024(July)-Bibtext_316_sample.txt

Here is the full Bibtex export

 FORMIS 2024(July)-Bibtex.txt

Here is the reformat of Bibtex into JSON.

 FORMIS_2024_July_Bibtex.json

The goals of this phase of FORMINDEX were to:

1. **Demonstrate the ability of analysis and visualization methods to scale to the full size of FORMIS** (~80,000 records) as well as to arbitrary subsets of citations (targeted bibliographies).
2. **Speculative and Realistically explore the use of synthetic intelligence** and augmented multisensory methods for Myrmecology.

Methods & Results

Methods & Results are separated into two main sections, reflecting the two goals above.

The first section "**FORMIS analysis**" describes the initial intra-FORMIS analysis methods performed & their location in the open source code.

The second section "**Generative AI methods**" describes three applications of Generative AI methods.

FORMIS analysis

Ingress and Target Bibliographies

- The [Read_in_FORMIS.py](#) script reads in the Bibtex and stores all records in JSON format.
- The [Generate_Target_Bibliographies.py](#) will generate subsets of all records, based upon the inclusion of key terms. Here there is an arbitrary indenting into 3 rows, so that the first row is a clade name, second row is Ant topics/terms, and third row is places.

```
TARGET_BIBLIOGRAPHIES = [  
    "Formica", "Camponotus", "Myrmica", "Pheidole", "Pogonomyrmex", "P. barbatus", "Monomorium", "Messor",  
    "Cataglyphis",  
    "Foraging", "Bioenergetics", "Myrmecophiles", "Ant-plant", "Dopamine", "Serotonin", "Art", "Robotics",  
    "Hawaii", "India", "Mexico", "Brazil", "California", "Intelligence"]
```

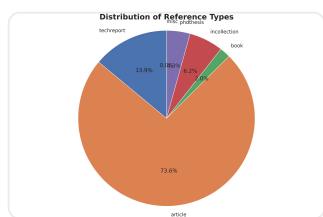
Targeted bibliographies are output to: https://github.com/docxology/FORMINDEX/tree/main/Targeted_Bibliographies

Visualization

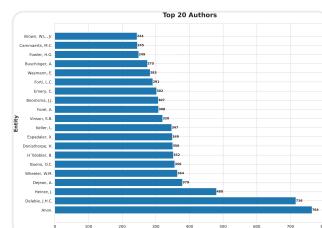
- [Visualize_FORMIS.py](#) is a set of visualization methods that apply to the full FORMIS JSON database, and any Target Bibliographies in the folder.
- The visualization outputs are sub-folders of <https://github.com/docxology/FORMINDEX/tree/main/Visualizations>.

Here is the
folder for all-FORMIS
visualizations

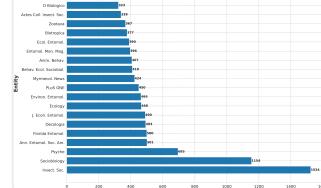
Distribution of record types

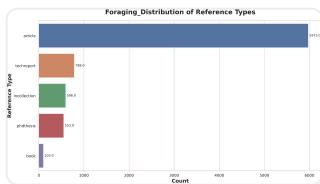


Top authors

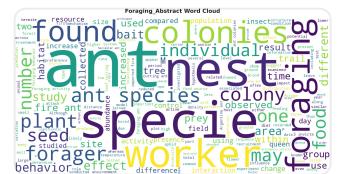
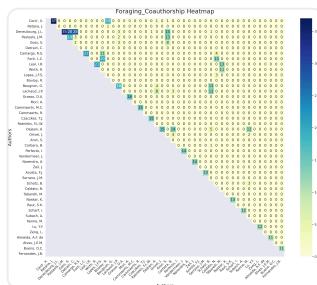
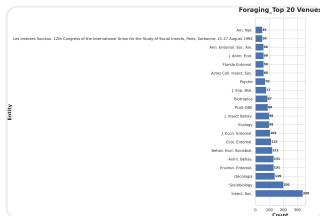


Top 20 Authors





Top publishing venues



Top topics and relevant records from term analysis.

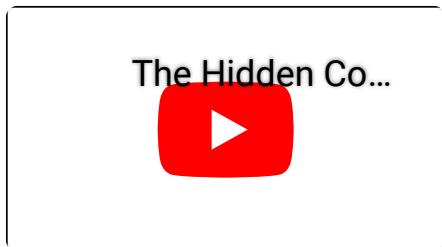
Generative AI methods

Three Generative AI methods were used: NotebookLM, OpenAI, and Perplexity.

NotebookLM

Google's NotebookLM was used <https://notebooklm.google.com/> to produce conversational podcasts on target sub-FORMIS-scale bibliographies. The target bibliography in JSON format was uploaded to a new Notebook. Then unless errors arose, I clicked "Generate Podcast". RevidAI <https://revid.ai/> was used on the audio files to add video & captions.

Several short (~10-20 minutes) podcasts for target sets of FORMIS citations, are hosted on my YouTube channel.



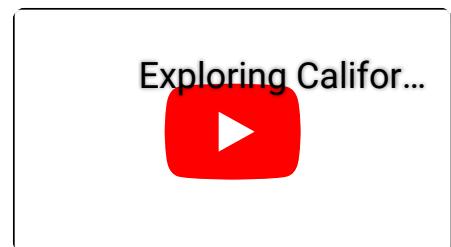
"Dopamine"

"Bioenergetics"

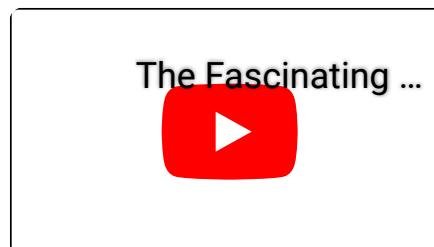


Red Harvester Ant ("P. barbatus")

"California"



"Intelligence"



OpenAI API

OpenAI LLM API was used for summarization and translation, all methods in https://github.com/docxology/FORMINDEX/tree/main/LLM_Methods

- Script 1 concatenates each of the target bibliographies with a general summarization prompt.
 - Script 2 uses an LLM call via OpenAI API to send the results of script 1 (the pro-summary) and receive the summary output.
 - Script 3 takes in literature summaries and translates into target non-English languages.

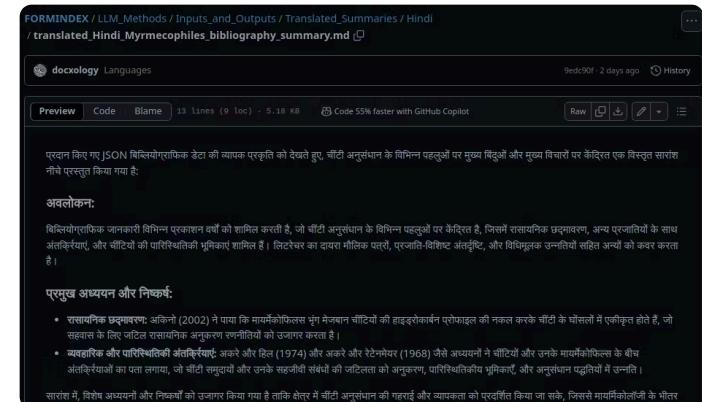
Outputs are all in

https://github.com/docxology/FORMINDEX/tree/main/LLM_Methods/Inputs_and_Outputs

The total LLM cost, using GPT-4o-mini, per summary & translation depends on the length of input and output, probably on the order of one-tenth to several cents USD in October 2024 (they are cheaper/different methods as well).

For example, here is the Myrmecophile targeted literature summary, translated into Hindi.

https://github.com/docxology/FORMINDEX/blob/main/LLM_Methods/Inputs_and_Outputs/Translated_Summaries/Hindi/translated_Hindi_Myrmecophiles_bibliography_summary.md



Perplexity API

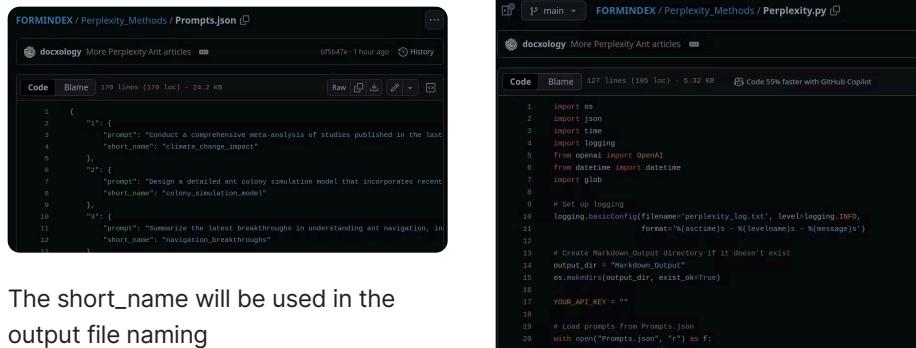
Perplexity.ai LLM API was used https://github.com/docxology/FORMINDEX/tree/main/Perplexity_Methods for internet-enabled Myrmecological augmented inquiries.

Prompts are listed here:

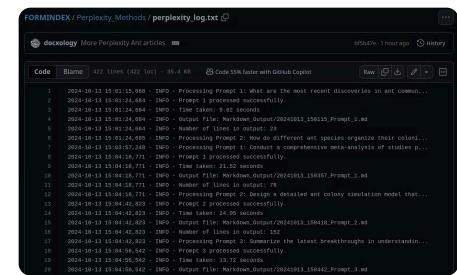
https://github.com/docxology/FORMIN-DEX/blob/main/Perplexity_Methods/Problems.json

[Perplexity.py](#) iterates over Prompts. For now, enter your Perplexity API in plaintext at line 17 in order to use the script.

The log captures the inputs, outputs, and timings. It takes about 10-40 seconds per prompt, and the number of returned lines is ~20-200



The short_name will be used in the output file naming



There is customizable system prompt



All outputs from Perplexity are in

https://github.com/docxology/FORMINDEX/tree/main/Perplexity_Methods/Markdown_Output

Just to give 3 examples:

Pogonomyrmex barbatus recent literature analysis

Career meta-analysis of Prof. Tim Linksvayer

Meta-analysis on recent Atta studies

Next steps

- Communicate with FORMIS stakeholders and see if there is any feedback or suggestions for useful analyses/visualizations.
- FORMIS analyses
 - Improve analysis and visualization on the FORMIS dataset.
 - Publish an open source report to describe the tools and analyses available at <https://github.com/docxology/FORMINDEX>. We can continue development at <https://github.com/activeInferenceInstitute/> for better long-term stewardship.
- Beyond FORMIS July 2024 snapshot.
 - Add more recent literature using literature engine APIs.
 - Integrate bibliographic records with NCBI species ID: <https://www.ncbi.nlm.nih.gov/taxonomy>. Later this will be a key identifier for the broader MetaInformAnt integrative effort.
 - Perplexity searches for all species groups.
 - Integration with fabric and other coordination mechanisms ([Active InferAnts Stream #004.1](#))

