#semanticClimate
Transforming information into actionable knowledge

BRIC
a DBT Organization

AI for Research

Automated Literature Retrieval

Presented by
Renu Kumari

Automated Literature Search?

Searching Scientific literatures for any query term from the publicly available databases and extracting information from the corpus using computational tools and techniques to reduce the time and effort required for traditional manual searches.

Why need ALR?

- Large no. of publications
- Time Efficiency
- covering a broader range of studies
- review stays up to date with the latest findings
- Consistency and Reproducibility
- Cost Efficiency

Uses of Literature Review

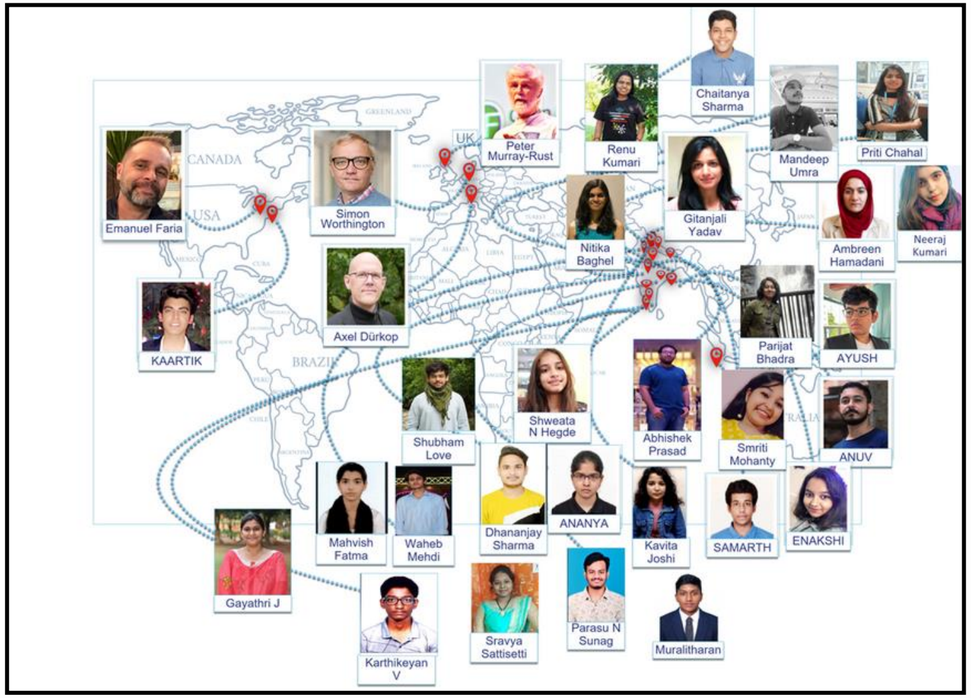
- Project grant writing
- Provides background information for a research problem
- Clinical Decision Support
- Identifies gaps in current knowledge.
- Defining Research Questions and Objectives

Tools has been developed by

What is #semanticClimate

A global citizen science movement using open notebook science and creating 5-star climate document

Team (Present and past)



Outreach

#semanticClimate
Transforming information into actionable knowledge

BRIC
a DBT Organization

Tools

1. **pygetpapers**
(https://github.com/petermr/pygetpapers)
-developed by Ayush Garg

Automated literature retrieval

2. **docanalysis**
(https://github.com/petermr/docanalysis)
-developed by Shweata N Hegde

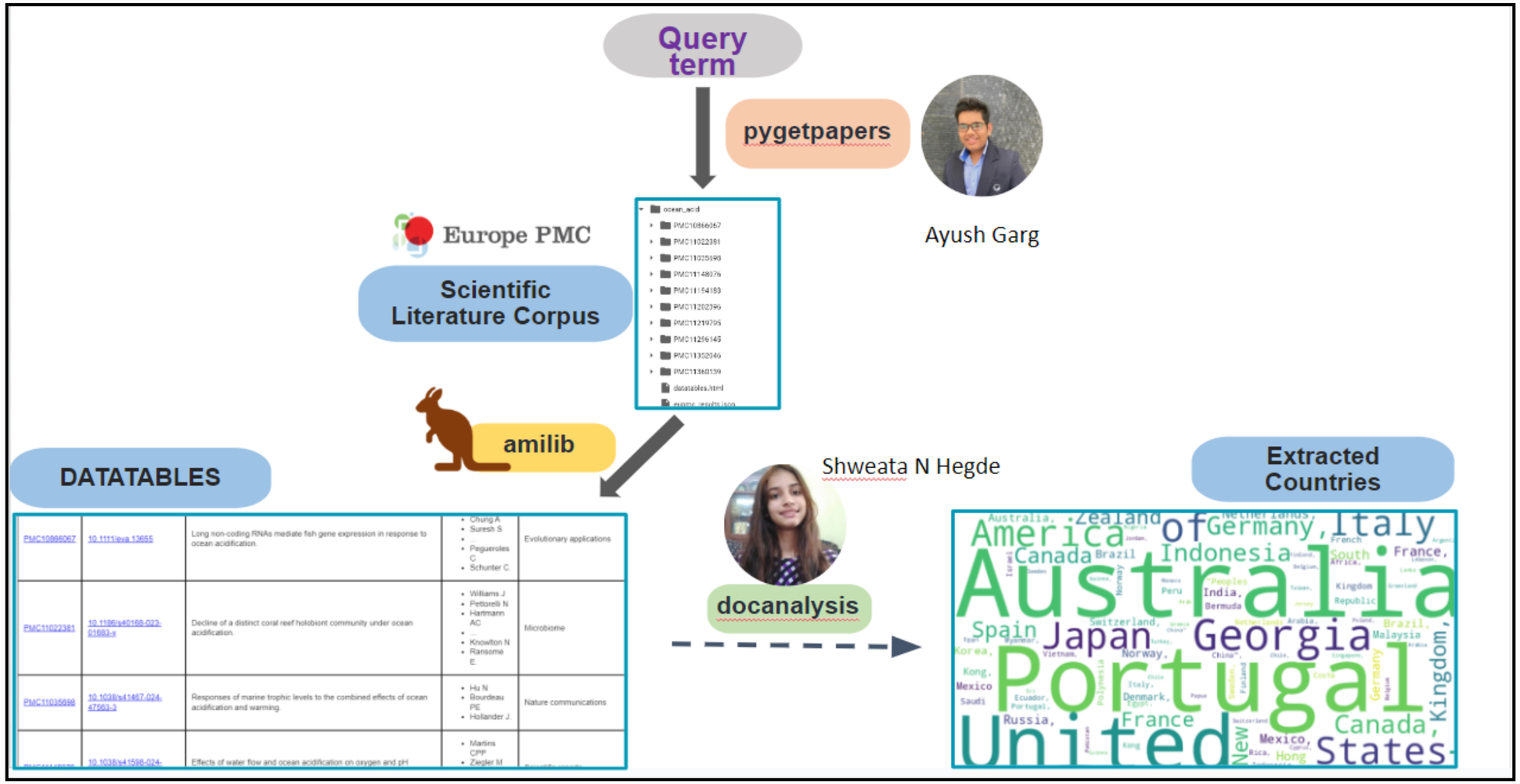
Entity search

3. **amilib**
(https://github.com/petermr/amilib)
-developed by Peter Murray-Rust

Create Datatable

Workflow

(Automated Literature Retrieval)



Google Colab (Collaboratory)

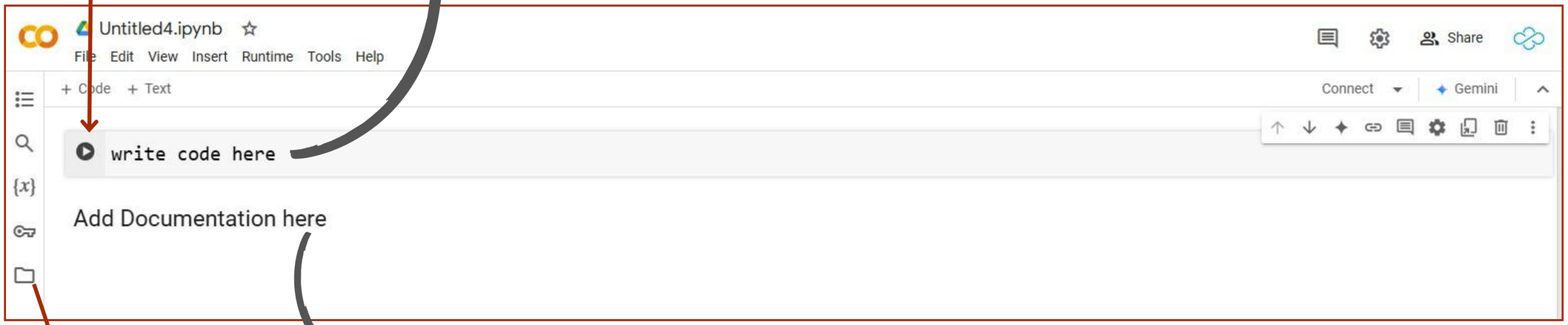
- free, cloud-based platform to write and execute python-based codes
- eliminates the need for complex installations
- access to powerful computing resources for machine learning and data analysis

Need Google account to get started!

Colab notebook

Click here to run the code

Code cell: for writing code



Text cell: for adding documentation

content folder: for saving output

ALR Workflow in Colab notebook

Colab Notebook URL: https://colab.research.google.com/drive/1Fg9DmgES0GroMmeB9_cBfCcDiR8uFSVH?usp=sharing

Tutorial URL : https://github.com/semanticClimate/bioAI_workshop/blob/main/Day1_ALR/tutorial_lit_search.md

Step 1: Tool Installation

```
[ ] !pip install --quiet pygetpapers
    !pip install --quiet amilib==0.3.9
    !pip install --quiet docanalysis
```

Step 2: Searching Scientific Literatures for any query

```
[ ] !pygetpapers --query '"phytochemicals" AND "Invasive plant"' --xml -n --output phytochem --save_query
INFO: Total number of hits for the query are 132
```

query

output dir

Step 3: Searching Lit for specific time frame

```
[ ] !pygetpapers --query '"phytochemicals" AND "Invasive plant"' --xml -n --startdate "2010-01-01" --enddate "2024-10-31" --output phyt

INFO: Total number of hits for the query are 123
```

Step 4: Adding limit to the number of papers to download metadata

```
[ ] !pygetpapers --query '"phytochemicals" AND "Invasive plant"' --xml --limit 100 --output phytochem --save_query

INFO: Total Hits are 132
WARNING: Could not find more papers
100it [00:00, 162759.18it/s]
INFO: Saving XML files to /content/phytochem/*/fulltext.xml
100% 100/100 [01:37<00:00,  1.03it/s]
```

Step 5: Summary table for retrieved scientific articles

```
[ ] !amilib HTML --operation DATATABLES --indir phytochem
```

Step 6: Display summary table

output

10 entries per page

Search:

query: "phytochemicals" AND "Invasive plant"; start: ...; end: 2025-01-08

pmcid	doi	title	authorString	journalInfo.journal.title	pubYear	abstractText
PMC10005159	10.3390/plants12051014	The First Evidence of Gibberellic Acid's Ability to Modulate Target Species' Sensitivity to Honeysuckle (Lonicera maackii) Allelochemicals.	• Barta CÉ • Jenkins BC • Lindstrom DS • Zahnd AK • Székely G.	Plants (Basel, Switzerland)	2023	Invasive species employ competitive strategies such as releasing allelopathic chemicals into the environment that negatively impact native species. Decomposing Amur honeysuckle (Lonicera maackii
PMC10072944	10.1098/rspb.2023.0055	Sunflower plantings reduce a common gut pathogen and increase queen production in common eastern bumblebee colonies.	• Malfi RL • McFrederick QS • Lozano G • Irwin RE • Adler LS.	Proceedings. Biological sciences	2023	Community diversity can reduce the prevalence and spread of disease, but certain species may play a disproportionate role in diluting or amplifying pathogens. Flowers act as both sources of nutrition

Step 7: Extracting list of Countries

```
!docanalysis --project_name phytochem --make_section --search_section INT,
RES, CON, DIS --dictionary COUNTRY --output phyto_invasive.csv
```

Step 8: Result- list of Countries

Step 9: Extracting list of Plants

```
[ ] !docanalysis --project_name phytochem --make_section --dictionary EO_PLANT --output phytoinvasive.csv --make_json phytoinvasive.jsc
```

Step 10: Result- list of Plants

Step 11: Extracting list of compounds

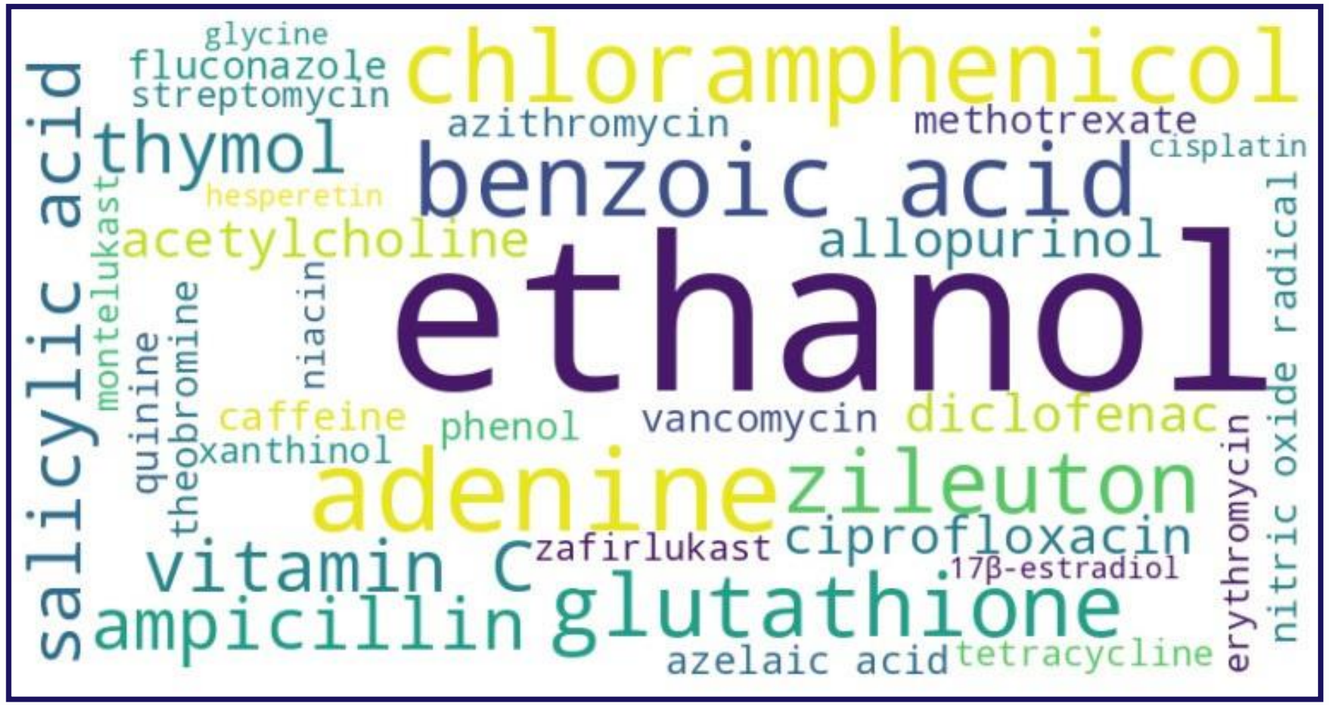
```
[ ] !docanalysis --project_name phytochem --make_section --search_section INT, RES, CON, DIS --dictionary EO_COMPOUND --output comp_
```

Step 12: Result- Chemical compounds

Step 13: Extracting list of drugs/medicine

```
[ ] !docanalysis --project_name phytochem --make_section --search_section INT, RES, CON, DIS --dictionary DRUG --output drug_invasive.
```

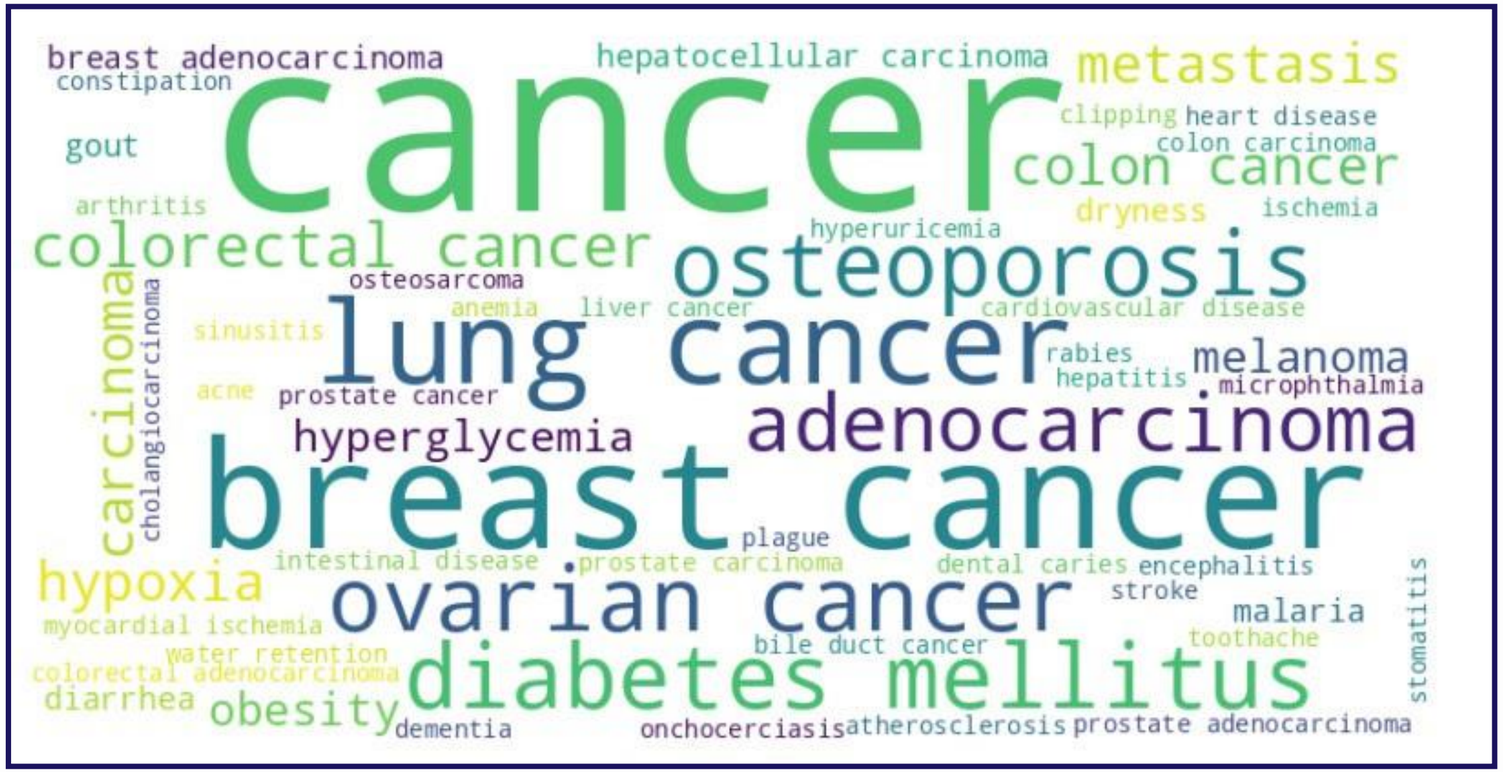
Step 14: Result- List of drugs



Step 15: Extracting list of Diseases mentioned

```
[ ] !docanalysis --project_name phytochem --make_section --search_section CON, DIS --dictionary DISEASE --output disease_all1.csv --ma
```

Step 16: Result- List of Diseases



Step 17: Extracting list of target organisms mentioned

```
[ ] !docanalysis --project_name phytochem --make_section --search_section CON, DIS --dictionary EO_TARGET --output target_info.csv --m
```

Step 18: Result- list of target organisms

#semanticClimate
Transforming information into actionable knowledge

BRIC a DBT Organization

NIPGR

Acknowledgements

Director, NIPGR
Dr. Gitanjali Yadav, Scientist, NIPGR
Prof. Peter Murray-Rust, Cambridge University
Simon Worthington, TIB, Germany
Ayush Garg
Shweata N Hegde
Lab 204
team #semanticClimate

Thank you

Contact us

Website : [https://semanticclimate.github.io/p/en/]
email : semanticclimate@gmail.com
X : [@semanticClimate]
LinkedIn : [@semantic Climate]
Git hub : [https://github.com/semanticclimate]