REPORT

#semanticClimate:
Developing a Climate dictionary from the IPCC Report: Food, Fibre, and Ecosystem Products

Submitted by:

ANMOL NEGI

SUMMER INTERNSHIP
MRIIRS, Faridabad

Under the Supervision of

Dr. Gitanjali Yadav
NIPGR, New Delhi

Dr. Peter Murray-Rust
University of Cambridge, UK

National Institute of Plant Genome Research
New Delhi 110067

August, 2024

Contents

List of Figures

1. Introduction

#semanticClimate is a citizen science effort using Open Notebook Science to transform information into structured, filtered, and actionable knowledge. Open Notebook Science involves sharing the entire research process in real-time, promoting transparency and collaborative efforts throughout the scientific investigation. Our mission is to liberate knowledge from climate-related reports and make it freely accessible.

The United Nations periodically publishes climate reports through several channels including the Intergovernmental Panel on Climate Change (IPCC) and United Nations Framework Convention on Climate Change (UNFCCC). It is important to realise that the availability of these reports to the public does not ensure their accessibility.

The focus is on two main activities: developing software tools for searching climate change literature and fostering community engagement through events like hackathons, datathons, data workshops and panel discussions.

The toolkit developed by #semanticClimate uses Artificial Intelligence (AI) over Natural Language Processing (NLP) to transform locked literature (PDFs) into semantic hypermedia (HTML/XML). The software tools include *pygetpapers*, which helps users find and download scientific articles, and *docanalysis*, which analyses these articles for important terms and information. These tools also support image processing, document conversion, and the creation of knowledge graphs and translations using resources like Wiki data and Wikimedia.

At present, our latest tools *amilib* and *amiclimate* are transitioning from their alpha to beta phase, which means the software is becoming ready for public use. This shift involves scaling community activities and adapting to new challenges, such as exploring UN Climate corpora and expanding citizen science events across India. We believe that openness is a powerful accelerant for finding solutions to our most pressing challenges.

2. Aims and Objectives

The IPCC prepares comprehensive Assessment Reports about knowledge on climate change, its causes, potential impacts and response options. It also produces Special Reports, which are an assessment on a specific issue and Methodology Reports, which provide practical guidelines for the preparation of greenhouse gas inventories.

Materials are published generally in three volumes, one for each of the Working Groups of the IPCC, plus a Synthesis Report. Each of the Working Group volumes is composed of individual chapters, an optional Technical Summary and a Summary for Policymakers.

- Working Group I examines the physical science underpinning past, present, and future climate change.
- Working Group II assesses the impacts, adaptation and vulnerabilities related to climate change.
- Working Group III focuses on climate change mitigation which involves actions that reduce the rate of climate change.

As a part of #semanticClimate, my tasks during the internship tenure included:

1. Exploring the toolkit used by the team, and understanding the cause that #semanticClimate is working towards. Going through the provided resources and documentation of the software developed in-house.

2. Participating in alpha testing and debugging sessions.

3. Overviewing the chapters of the Sixth Assessment Report (AR6) and selecting one to contribute to, by creating a wordlist for the comprehensive semantic Dictionary that is currently in development.

I undertook work on Chapter 5, of Working Group II - "**Food, Fiber and Other Ecosystem Products**". Below is a summary of the chapter -

*The **IPCC AR6 Working Group II Chapter 5** examines the cascading impacts of climate change on **food, fiber, and ecosystems**, highlighting significant threats to food security and biodiversity. Rising temperatures, extreme weather events, and shifting climate patterns are already disrupting agricultural productivity, fisheries, and livestock systems, with disproportionate effects on vulnerable, low-income regions. Biodiversity loss and ecosystem degradation further threaten essential services, intensifying risks to livelihoods and food supply chains. Adaptation measures like climate-resilient crops, agroforestry, and sustainable fisheries management are crucial, but uneven implementation and funding gaps hinder progress. The report stresses the interdependence of food systems and ecosystems, calling for integrated management and global cooperation. Transformative actions such as reducing food waste, transitioning to sustainable diets, and enhancing local resilience are vital for mitigation. Without rapid mitigation, food systems risk irreversible damage, exacerbating inequality and resource scarcity. Strengthened global commitments, financial investments, and knowledge-sharing are essential to implement climate-resilient development pathways. These actions are critical to safeguarding food security, reducing inequality, and preserving ecosystem integrity in the face of escalating climate challenges.*

4. Working on a section of the IPCC Glossary, by adding Wiki data and Wikipedia links. Contributed author in document of Ranganathan2024 for ALIS publication.

3. Tools and Technologies

Tools used to accomplish the aforementioned objectives include, but are not limited to:

- *pygetpapers:* A valuable text mining tool that interacts with open access scientific repositories, systematically acquiring relevant articles.

- *docanalysis:* An integrated suite of open-source Command Line tools that enables users to download scientific literature from *europepmc.org* based on specific query criteria.

- *amilib:* A Python library that facilitates the conversion of PDFs into semantic

HTML. A non-exhaustive list of software and technologies critical to our functioning:

- Git: A distributed version control system that tracks versions of files.

- GitHub: A developer platform used extensively for hosting our code. It is accessible to the public and easy to collaborate on.

- Python: A powerful high-level programming language that is ideal for scripting and rapid application development.

- VS Code or PyCharm: A highly efficient Integrated Development Environment (IDE) that allows coding, testing and debugging for software development.

- Google Colab: A hosted Jupyter Notebook service that requires no setup to use and provides free of charge access to computing resources, including GPUs and TPUs.

- Google Workspace: A feature-rich range of tools for building and maintaining projects, such as Google Docs, Google Sheets etc.

- Slack: A cloud-based team platform for seamless communication and coordination.

4. Deliverables

Link to the chapter summary in Markdown:
https://github.com/petermr/amilib/blob/anmol_test/temp/words/Food_Ecosystem_Executive_summary.md

Link to the curated wordlist of more than 150 words:
https://github.com/petermr/amilib/blob/anmol_test/test/resources/wordlists/food_ecosystem.txt

Link to Dictionary in semantic HTML:
https://github.com/petermr/amilib/blob/anmol_test/temp/dictionary/climate/food_ecosystem.html

Link to Dictionary in XML format:
https://github.com/petermr/amilib/blob/anmol_test/temp/dictionary/climate/food_security.xml

4.1 Snapshots

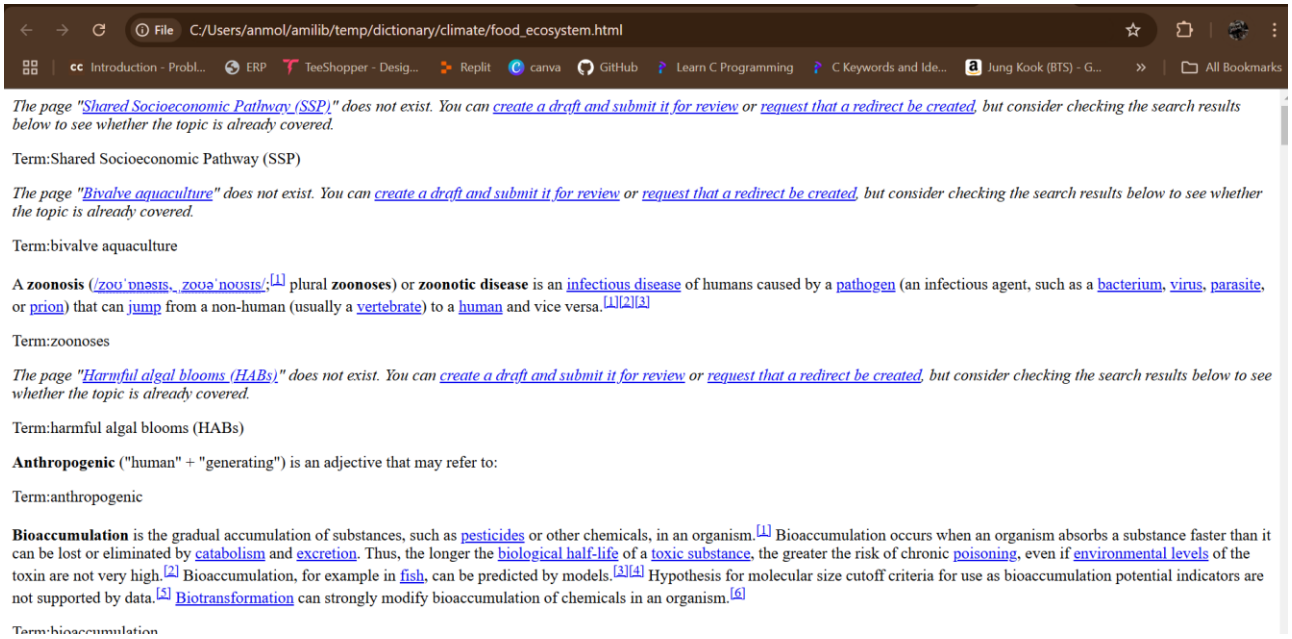

Figure i.

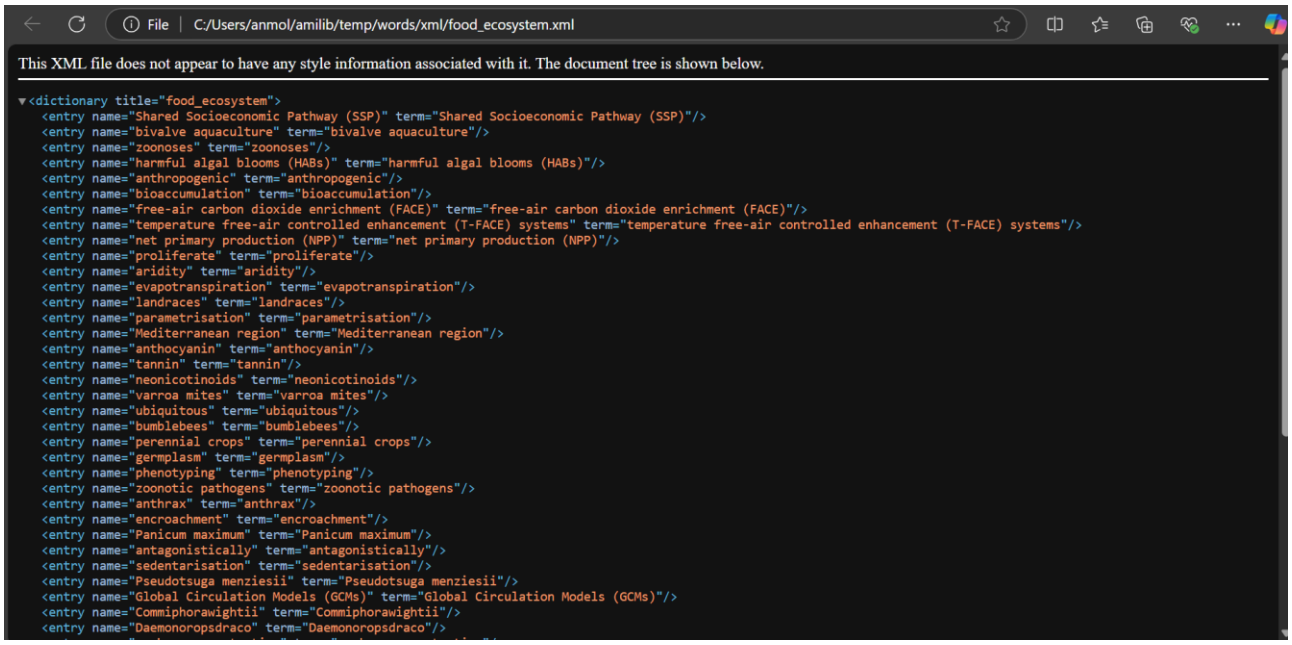

Figure ii.

Figure iii.

5. Conclusion

During the internship, I extensively used Git as a version control system. This helped me get skilled in managing code changes, collaborating with team members, and maintaining a clean and efficient codebase. I was actively involved in daily testing and debugging sessions, which included running and creating unit tests for our open-source tool, amilib.

I had the opportunity to delve into the world of semantics and Tim Berners-Lee's semantic web. Understanding how to represent data meaningfully and interconnect information across different domains was enlightening and expanded my knowledge in this area. Additionally, I participated in paper drafting sessions for the Annals of Library and Information Sciences (ALIS) Journal, which provided valuable insights into the history of semantic publishing. These sessions highlighted the crucial need for open source and open access in academic and scientific research, emphasizing the importance of making knowledge freely available.

My hands-on experience with Python fundamentals in this 8-week period also gave me exposure to the numerous scientific applications of the popular high-level language. As a Summer Research Fellow of the Indian Academy of Sciences, I was required to send regular work update reports to the Academy, which made me learn and adopt effective documentation practices.

Overall, this internship was a great learning experience. It provided me with practical skills in version control, testing, documentation, semantic and open-source technologies, and Python programming.