REPORT

#semanticClimate:
Bridging the Gap between Data Availability and Accessibility

Submitted by:

Nitika Baghel

in partial fulfilment of the requirements of
PC649 SUMMER INTERNSHIP
DA-IICT, Gandhinagar

Under the Supervision of

Dr. Gitanjali Yadav   Dr. Peter Murray-
                                    Rust
NIPGR, New Delhi   University of Cambridge, UK

National Institute of Plant Genome Research
New Delhi 110067

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 Wikidata 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 8, of Working Group II - "**Poverty, Livelihood and Sustainable Development**". Below is a summary of the chapter -

Climate change significantly affects the poorest and most vulnerable, worsening poverty and livelihoods. This chapter examines how climate change impacts people, focusing on vulnerability, poverty, and livelihoods. It highlights the importance of climate justice and historical inequalities in shaping responses. Past reports have shown that disadvantaged groups face the worst impacts. To adapt, societies need equitable development, robust institutions, and poverty reduction. The chapter uses global and local data to assess human vulnerability and losses from climate change, emphasising both economic and non-economic impacts. It explores how climate change affects livelihoods and resilience, stressing the need for integrated strategies for sustainable development and adaptation. Climate justice, legal, and institutional frameworks are crucial for effective adaptation.

4. Working on a section of the IPCC Glossary, by adding Wikidata and Wikipedia links, and also adding Indian language equivalents (Gujarati, in my case) to the existing terms.

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:
[amilib/temp/misc/wg2_chap8_summary.md at nitika_test · petermr/amilib (github.com)](#)

Link to the curated wordlist of more than 150 words:
[amilib/test/resources/wordlists/wg2_chap8.txt at nitika_test · petermr/amilib (github.com)](#)

Link to Dictionary in semantic HTML:
[amilib/temp/words/html/wg2_chap8.html at nitika_test · petermr/amilib (github.com)](#)

Link to Dictionary in XML format:
[amilib/temp/words/xml/wg2_chap8.xml at nitika_test · petermr/amilib (github.com)](#)

4.1 Snapshots

Figure i.

This XML file does not appear to have any style information associated with it. The document tree is shown below.

▼<dictionary title="wg2_chap8">
  <entry name="Poverty" term="Poverty"/>
  <entry name="Livelihood" term="Livelihood"/>
  <entry name="Sustainable Development" term="Sustainable Development"/>
  <entry name="atoll" term="atoll"/>
  <entry name="pandemic" term="pandemic"/>
  <entry name="Spatial hotspot" term="Spatial hotspot"/>
  <entry name="equity" term="equity"/>
  <entry name="equality" term="equality"/>
  <entry name="Economic losses" term="Economic losses"/>
  <entry name="Non-economic losses" term="Non-economic losses"/>
  <entry name="Response capacity" term="Response capacity"/>
  <entry name="socioeconomic" term="socioeconomic"/>
  <entry name="Paris agreement" term="Paris agreement"/>
  <entry name="synergy" term="synergy"/>
  <entry name="Trade-off" term="Trade-off"/>
  <entry name="Adaptation" term="Adaptation"/>
  <entry name="Mitigation" term="Mitigation"/>
  <entry name="Climate Resilience" term="Climate Resilience"/>
  <entry name="Adaptation finance" term="Adaptation finance"/>
  <entry name="pro" term="pro"/>
  <entry name="global trend" term="global trend"/>
  <entry name="shock" term="shock"/>
  <entry name="Urbanisation" term="Urbanisation"/>
  <entry name="Economic globalisation" term="Economic globalisation"/>
  <entry name="COVID-19" term="COVID-19"/>
  <entry name="Development deficit" term="Development deficit"/>
  <entry name="exacerbate" term="exacerbate"/>
  <entry name="human vulnerability" term="human vulnerability"/>
  <entry name="high confidence" term="high confidence"/>
  <entry name="resilient" term="resilient"/>
  <entry name="emission" term="emission"/>
  <entry name="adaptive responses" term="adaptive responses"/>
  <entry name="Global warming" term="Global warming"/>
  <entry name="onset" term="onset"/>
  <entry name="poverty trap" term="poverty trap"/>
  <entry name="food insecurity" term="food insecurity"/>
  <entry name="societal" term="societal"/>
  <entry name="impact" term="impact"/>
  <entry name="climate change" term="climate change"/>
  <entry name="mortality" term="mortality"/>
  <entry name="low middle income countries" term="low middle income countries"/>
  <entry name="disproportionally" term="disproportionally"/>
  <entry name="hazard" term="hazard"/>
  <entry name="migration" term="migration"/>
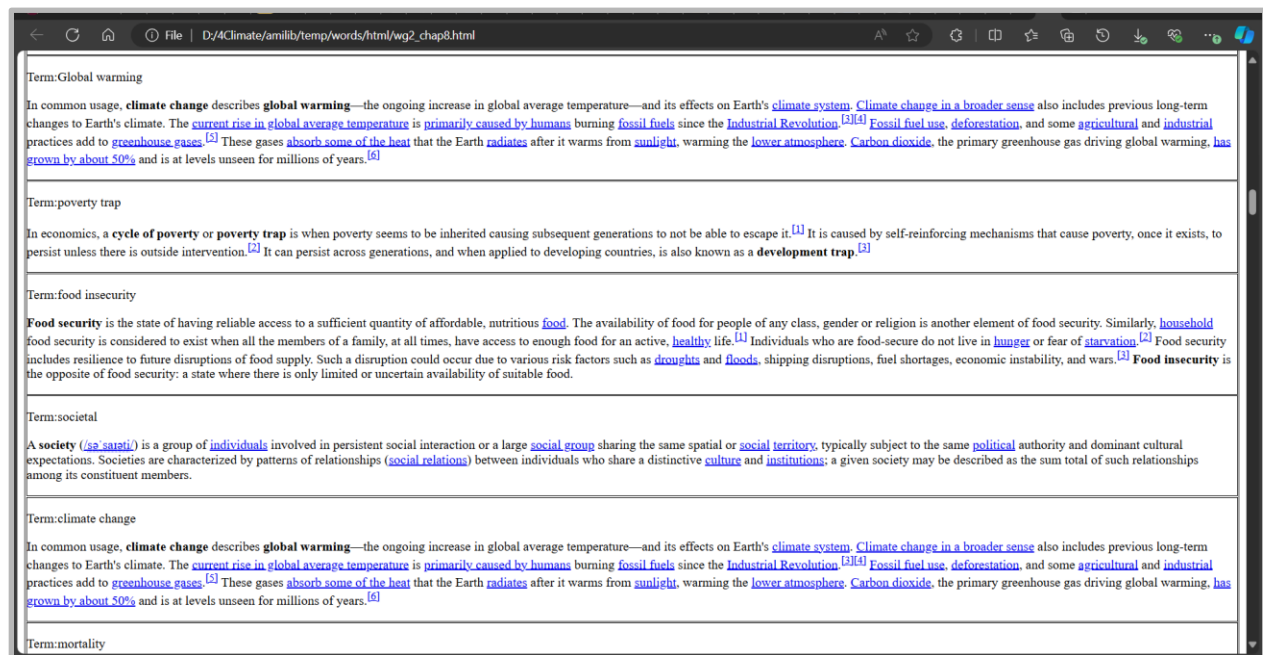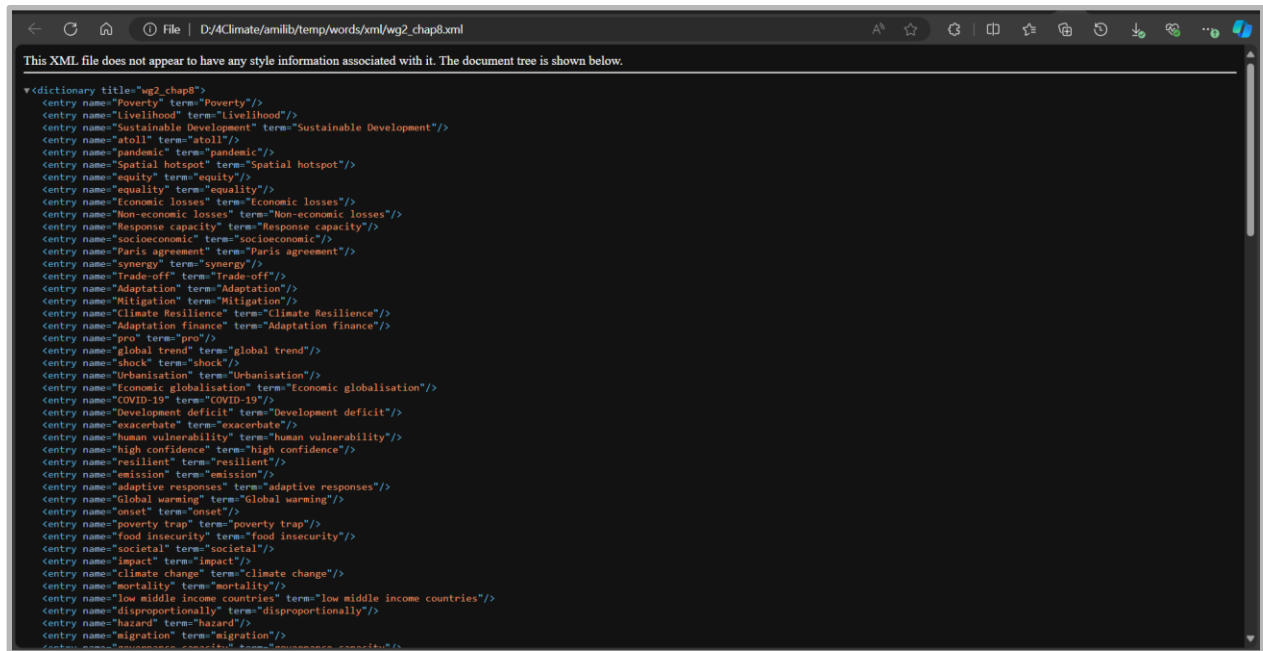 <entry name="governance capacity" term="governance capacity"/>

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, emphasising 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.