Novelty Components of Scientific Productions for Each SDG
-
Dashboard Project

Faculté
des **sciences économiques**
et de **gestion**
Université de Strasbourg
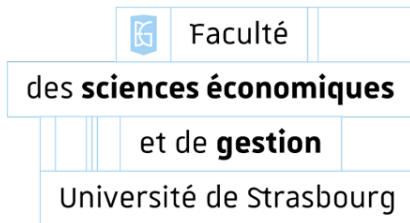
Supervised by: Diletta Abbonato
Authors: Valentin Barthel, Adrien Busché, Diego Fracassi
valentin.barthel@etu.unistra.fr
adrien.busche@etu.unistra.fr
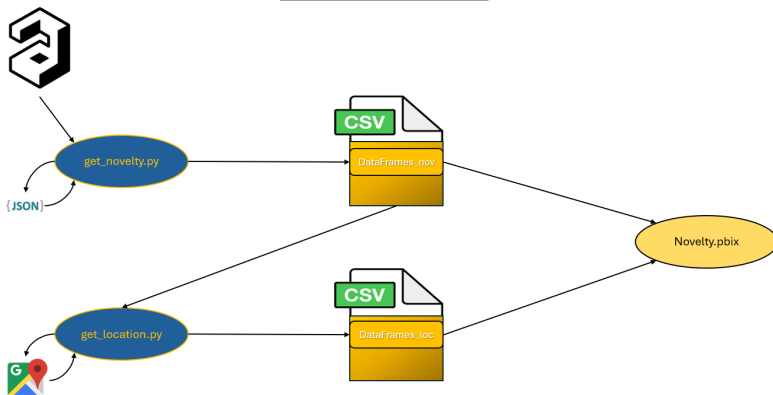diego.fracassi@etu.unistra.fr

Table of Contents

Inspiration

- Our idea is to draw inspiration from the work of Lee, Walsh and Wang (2015) to highlight insights into the components of the novelty potential of scientific outputs.
- We therefore calculate novelty scores based on the Commonness indicator for scientific outputs discussing the 17 (Sustainable Development Goals) SDGs independently of one another and check differences between them.
- We worked on articles listed in the OpenAlex database and used the Novelpy package (Pelletier and Wirtz (2022)).

Structure of the Project

Arborescence of the Project :

```
C:.
 get_location.py
 get_novelty.py
 main.py
 Novelty.pbix
 README.md
 requirements.txt
DataFrames_loc

 ...
DataFrames_nov

 ...
Presentation
NoveltySchema.png
Novelty_Support.pdf
```

Introduction
○○●

Novelty Indicator
○

Geocoding
○

Our Dashboard Structure
○

Limitations
○○

References

Our Program Diagram

$\Longrightarrow Github$

Co-occurrence Matrix

- The co-occurrence matrix represents how frequently two references (journals) appear together in a scientific article.
- Each cell (i, j) contains the number of times journal i and journal j are cited together.

Reference	Journal A	Journal B	Journal C	Journal D
Journal A	10	3	2	0
Journal B	3	15	1	1
Journal C	2	1	12	0
Journal D	0	1	0	8

Table: Example of a co-occurrence matrix

$\implies get\_novelty.py$

Geocoding

The aim is to identify the geographical location of institutions linked to authors of scientific publications.
Functions:

- **get_city_state(*place_name*)**: Fetches geographic details (city, region, state, latitude, longitude) for a given institution name (Google maps API).

- **get_city_from_coordinates(*lat, lng*)**: Finds city, region, and state from latitude and longitude (Google maps API).

- **get_continent_from_country(*country_name*)**: Determines the continent for a given country (pycountry_convert).

$\implies$ *get_location.py*

Our Dashboard Structure

This dashboard allows you to explore the novelty indicators for scientific productions on the 17 SDGs since 2016. It consists of 5 pages providing a complete analysis of the current and past situation.

- **'Summary'**: This page provides information about a number of SDGs. We can see the number of citations, publications, and novelty indicator since 2016 for each author, area of expertise, type of production, and the evolution by year.
- **'Year View'**: A year-by-year analysis of each SDG. We can see changes in key indicators compared with the previous year, as well as the most active SDGs by type and institution.
- **'Mapping'**: This gives you a geographical view of the scientific output situation.
- **'SDG Comparison'**: Allows you to compare the key indicators for different SDGs. We offer a view by year, by region, and analyze the differences in the link between the number of contributors and the novelty indicator.
- **'Database'**: A page that allows the user to check the data by being able to sort it.

Limitations of the Novelty Indicator

- **Dependence on citation patterns:** The indicator relies heavily on the availability and accuracy of citation data.
- **Field-specific dynamics:** It may not account for differences in citation practices across disciplines.
- **Computational limitations:** In very large datasets, constructing co-occurrence matrices and calculating frequencies can be resource-intensive.
- **Simplistic measure of novelty:** The indicator may not capture the full spectrum of creativity or innovation in the research being analyzed.

Our Limitations

- **Keyword Queries:**
    - The use of multiple queries per SDG is limited due to error 403 restrictions.
    - The selected queries may not necessarily capture the most cited scientific production for the thematic area.

- **Computational and Sample Size Limitations:**
    - Each query is constrained by the frequency of the selected keyword's occurrence in scientific papers.
    - The available RAM (16 GB) is insufficient for the demands of a comprehensive scientometric approach.

Introduction
000

Novelty Indicator
o

Geocoding
o

Our Dashboard Structure
o

Limitations
oo

References

Lee, You-Na, John P Walsh, and Jian Wang, "Creativity in scientific teams: Unpacking novelty and impact," *Research policy*, 2015, *44* (3), 684–697.

Pelletier, Pierre and Kevin Wirtz, "Novelpy: A python package to measure novelty and disruptiveness of bibliometric and patent data," *arXiv preprint arXiv:2211.10346*, 2022.