International health data and information discovery for Native/Indigenous peoples using web scraping

Project Title	International health data and information discovery for Native/Indigenous peoples using web scraping
Project Summary	Finding current and historical data on native/indigenous people can be a challenge. This project will use and refine automated search capabilities using your coding and algorithm skills, ideally in Python/Beautiful Soup, ArcGIS, or similar tools such as webscraper.io or Maptive.
Country	United States

Project Description

There are about 180 member states of the United Nations. Possibly, 20 of those states collect data on native/indigenous people in their territory. Indigenous nations (N=3,000 worldwide?) do not get the same attention internationally as member states do, and they fall into the margins of the data collected, if at all. In a world where borders are continually breaking down, there is a real need to know how indigenous people are faring in their health, livelihood, language, and culture. The raw material for further analysis is needed, and search, web scraping, and literature review can deliver it affordably.

Required Skills or Interests

Skill(s)
Coding
Data analysis
GIS expertise
Research
Software development
Website design

Additional Information

If your knowledge of code is more significant than your knowledge of languages, you are not alone. The ability to write code is more needed now more than ever due to the existence of international internet media. There are at least four forms of media (journal articles, data caches, unscanned paper monographs, and websites) where reliable data are known to exist about native/indigenous people, and only one of them is web-based, but it is the most accessible to reach. Also, there exist APIs that collect multiple snapshots of the same page as it evolves. The problem is that the actual use of web scraping or geographic information systems is not easy. Most of the internship will involve test cases where we see how to adapt the tools we have. It would be good if you knew another language than English, but thanks to Systran and Google, that's optional. What is needed is a willingness to use algorithms and code to discover in partnership what is there, a combination of the informationist and data scientist approaches to the wealth of data present in both caches and unstructured textbases. Exposure to another human language is helpful but not necessary.

Language Requirements

Language	Speaking Proficiency	Reading Proficiency	Importance
Arabic	Elementary proficiency	Limited working proficiency	Nice to Have
French	Elementary proficiency	Limited working proficiency	Nice to Have
Spanish	Elementary proficiency	Limited working proficiency	Nice to Have
Russian	Elementary proficiency	Limited working proficiency	Nice to Have
Chinese-Mandarin	Elementary proficiency	Limited working proficiency	Nice to Have
Turkish	Elementary proficiency	Limited working proficiency	Nice to Have
Uzbek	Elementary proficiency	Limited working proficiency	Nice to Have