

Capstone Project Proposal

Problem Statement: Which rhetorical, deliberative, and argumentative modes do NGOs employ to attempt to impact the way in which the UN Human, Social and Economic Rights council frames of and addresses environmental, developmental, and defense/foreign policy matters and guidelines? To what extent do these NGOs successfully shift these arguments/meanings? How do these differences in effectiveness/strategies vary over time, geographic location (e.g. African NGOs in the 1980s vs Western NGOs in the 2010s), and other demographic and cultural variables (e.g. Women's vs Indigenous vs Peace mvmts etc.)? Could also add comparisons w/ other intl organizations like UNESCO or World Bank if time

Data: text corpus of various NGO documents from UN Library – stratified sample to be representative (by geographical place + time + category of mvmt)—through an API

Decision Maker: UN administrative staff—how they would manage interventions

Other stakeholders: various NGOs

Success for this project: illustrating how these frames shift over time + identify the mechanisms through which NGOs attempt to shift meaning – improving complaint mechanism

Constraints: limited to document corpus and standard Python NLP programming techniques

Scope + Risks: UN could use these strategies to shut down project

Techniques to be used: Topic Analysis, Time Series Analysis

Find linkages over time periods to create river network

1. Problem identification: see above
2. Data wrangling:
Stratified randomization – select distribution of data from different subgroups and time periods of our various organizations
3. Exploratory data analysis:
4. Pre-processing and training data development:
Text chunking algorithm to identify noun sequences (could use n-grams to be a bit easier)
lemmatization
Grouping coherent terms using community detection algorithm to identify discursive patterns
Use Parts of Speech tagger like Treetagger
5. Modeling:
Named Entity Recognition—use PoS tagging to identify relevant noun phrases (Cointet)

Use discriminating words, checking differences in proportion to identify words used by particular actors (Nelson)

Downstream Structural Topic Models with covariates as geographical location, organization, time period--- can use topics to help predict covariates (Karell and Freedman)

Robustness checks: regression—check data is coherent

6. Documentation:

Time series of terms—how cultural valencies within pairs of associations change over time and between various places/countries (Jia)

Dynamic river networks—how differing topics become interlinked and transformed (Cointet and Bearman)

Global semantic network/web of all terms

Local semantic networks of certain time periods (could I also use certain spatial locations, like Africa vs Europe?)