Mapping Public Awareness
Dan C. Baciu, University of California, Santa Barbara

Project description:
This project uses text mining to analyze news. We are mapping geographical distributions and studying historical change. For example, we ask: Do news show where and when environmental awareness evolved in the Arctic? To answer such questions, we collect news and structure this text data using NLP tools, in particular with the Illinois Wikifier, and then analyze it using supervised as well as unsupervised approaches. Finally the results are mapped and studied through interactive interfaces previously developed in our lab. This term, we are getting started with two new datasets (Tasks 1 & 2). We are also streaming existent tools (Task 3).

Skills:
Python, NLP, Json format. Other skills: command line running Java; html, Javascript, D3, content management systems.

Task 1: Environmental Awareness in the Arctic
Svabard is the northernmost permanently inhabited Arctic archipelago. We evaluate the history of environmental awareness on these islands by mining local news as well as scholarly articles in English, Russian, and Norwegian. The results go into a research article by Abramova and Baciu see link below.
https://aag.secure-abstracts.com/AAG%20Annual%20Meeting%202020/abstracts-gallery/36757

Procedure: Collecting additional data, cleaning the data, wikifying. Classifying the data into texts that deal with local environmental issues and texts that deal with remote environmental issues. Mapping areas covered by both classes of articles over time.

Task 2: Archaeology in Belize
Belize has many archeological sites that are presently being explored, in particular those of the Mayan culture. We study scholarly articles written in the US. We map geographical locations mentioned in these articles, and we use dimensionality reduction to study the relationships between various locations.

Procedure: Collecting additional data, cleaning the text data, wikifying, mapping based on wikipedia, mapping using gazetteers, analysis through dimensionality reduction. Mapping of text sources (journal locations within the US). Visualization with our GeoD.

Task 3: Streamlining existing code
We developed a toolkit for text mining that integrates wikification with LSA and LDA and semantic testing for performance improvement, as well as interactive visualization. The code requires streamlining and documentation.

Procedure: Rewriting and documenting code. Running existent and new code on existent dataset.