Team Project Proposal

Topic: Reproducing a Paper: Mining causal topics in text data: Iterative topic modeling with time series feedback.

Team: PYM

First	**Last**	**email**
Pallavi	Ravada	pravada2@illinois.edu
Yash	Skhwani	yashas2@illinois.edu
Michael	Bernardoni	mlb12@illinois.edu

Captain: Michael   Bernardoni   mlb12@illinois.edu

GitHub fork link: https://github.com/mlbernardoni/CourseProject

Reproduce the following paper:

- Subtopic: Causal topic modeling
  - Hyun Duk Kim, Malu Castellanos, Meichun Hsu, ChengXiang Zhai, Thomas Rietz, and Daniel Diermeier. 2013. Mining causal topics in text data: Iterative topic modeling with time series feedback. In Proceedings of the 22nd ACM international conference on information & knowledge management (CIKM 2013). ACM, New York, NY, USA, 885-890. DOI=10.1145/2505515.2505612

Programming language:

The team will initially proceed programming with Python. However if performance becomes an issue C++ will be evaluated.

Tableau may be used for data visualization of the resulting datasets.

Dataset:

The dataset can be found at the Linguistic Data Consortium (https://www.ldc.upenn.edu). The team captain (mlb12@illinois.edu) has applied to the consortium and is await approval from: University of Illinois at Urbana-Champaign, Beckman Institute. This request has also been posted on Piazza.

Once approval is obtained access to the dataset will be provided.