COLERIDGE INITIATIVE

RICH CONTEXT COMPETITION

We are pleased to invite you to compete in a "Rich Context" text competition. The goal of this competition is to automate the discovery of research datasets in social science publications — so that analysts can see **what** other research has been done with the datasets, by **whom** and with what **results**. Participants can use any combination of machine learning and data analysis methods to find datasets and associated methods and research fields in a corpus of social science publications.

PRIZE

Finalists will be awarded a prize of $2,000 each. A stipend of $20,000 will be awarded to the winning team; the winning team will work with the sponsors in the subsequent implementation of the algorithm.

STRUCTURE

The competition has two phases. In the first phase, you will be provided labeled data, consisting of a corpus of 2,500 publications matched to the datasets cited within them. You can use this data to train and tune your algorithms. In the second phase, you will be provided with a large corpus of unlabeled documents and asked to identify the datasets used in the documents in a test corpus, as well as the associated methods and research fields. You will be scored on the accuracy of your techniques, the quality of your documentation and code, and the efficiency of the algorithm– and also on your ability to find methods and research fields in the associated passage retrieval.

TIMELINE

September 30, 2018: Participants submit a letter of intent (see How to Participate)

October 15, 2018: Participants notified and Phase 1 data provided (see First Phase Participation)

November 15, 2018: Preliminary algorithm submitted (see Program Requirements)

December 1, 2018: 15 finalists selected (see First Phase Evaluation) and Phase 2 data provided (see Second Phase Participation)

January 15, 2019: The algorithms of up to 6 teams are selected for final submission (see Second Phase Evaluation)

February 15, 2019: Workshop is held in New York for final presentation and selection of winning algorithms (see Second Phase Evaluation)

ADDITIONAL INFORMATION

Competition Website: coleridgeinitiative.org/richcontextcompetition

All submitted algorithms will be made publicly available as open source tools.