

Tailoring Named Entity Recognition (NER) to extract pest event data from online news and Tweets

Ariel Saffer, Laura Tateosian, Makiko Shukunobe, Chelsey Walden-Schreiner, Center for Geospatial Analytics, NC State University

Plant pest and pathogen (pests, for brevity) observation and distribution data for biosecurity measures and forecasting models commonly come from field observations, official reports, and genetic records consolidated through published literature and databases. While vital, these data can suffer from latency and spatial and temporal sparseness, due to the cost of collecting and collating these data at scale. These limitations have implications for the predictive capabilities of models and the success of control and eradication programs. At the same time, there is growing evidence that Web media, such as online news and Twitter, could be a valuable source of supplementary data to boost coverage and tap into up-to-date information for predicting pest encroachments. We present our results demonstrating the value of data from web media to supplement official records of pest invasions, using Spotted Lanternfly and *Tuta absoluta* as case study species. Our ongoing research explores and evaluates text analysis and Natural Language Processing approaches like Named Entity Recognition (NER) for automating the extraction of structured data from these unstructured sources, with potential for both historical and real-time applications.