The Journal of Open Source Software

PETRARCH2: Another Event Coding Program

Clayton Norris[1], **Philip Schrodt**[2], **and John Beieler**[3]

1 University of Chicago **2** Parus Analytics **3** Human Language Technology Center of Excellence, Johns Hopkins University

DOI: 10.21105/joss.00133

Software

- Review
- Repository
- Archive

Summary

The PETRARCH2 coding program implements a new coding algorithm, based on a syntactic constituency parse, to extract who-did-what-to-whom political event data from structured news stories. Events are coded according to the CAMEO (Gerner et al. 2001) coding ontology. This software improves upon previous-generation coding software such as TABARI (Schrodt 2001) by using a deep syntactic parse rather than shallow parsing.

At the level of assigning codes, PETRARCH2 is largely dictionary based, working from extensive dictionaries of verb phrases to identify the type of event, and noun phrases to identify both the actor (generally a proper noun such as the name of a country or leader) and agent (generally a common noun identifying a role such as "police" or "protesters"). These dictionaries incorporate the synonym sets from WordNet, are open source, and are included in the distribution.

PETRARCH2 has primarily been run using Treebank output from the Stanford CoreNLP system. It can be integrated with other software on the https://github.com/openeventdata/ site to handle either continuous near-real-time coding or batch coding, as well as auxiliary programs for geolocation and simple deduplication.

As noted above, PETRARCH2 extracts structured political events from news stories. This software is aimed at users in the social sciences for whom the extracted political events can serve as basis for research into phenomenon of interest. Additionally, other users, such as data journalists, data scientists, and other analysts, can make use of the data generated via PETRARCH2 as a structured, common representation of news stories.

References

Gerner, Deborah J., Philip A. Schrodt, Omur Yilmaz, and Rajaa Abu-Jabr. 2001. "Conflict and Mediation Event Observations (Cameo): A New Event Data Framework for the Analysis of Foreign Policy Interactions."

Schrodt, Philip A. 2001. "Automated Coding of International Event Data Using Sparse Parsing Techniques."