Template for sri.com publication posting **- PA-22-015**

**Title**

*Accelerating Human Authorship of Information Extraction Rules*

**SRI Authors**

*Dayne Freitag, John Niekrasz, John Cadigan, Robert Sasseen*

**Publication month, year**

*October, 2022*

**Citation**

Dayne Freitag, John Cadigan, John Niekrasz and Robert Sasseen. 2022. “Accelerating Human Authorship of Information Extraction Rules.” Proceedings of the *First* *Workshop on Pattern-Based Approaches to NLP in the Age of Deep Learning* (PAN-DL).

**Abstract**

We consider whether machine models can facilitate the human development of rule sets for information extraction. Arguing that rule-based methods possess a speed advantage in the early development of new extraction capabilities,

we ask whether this advantage can be increased further through the machine facilitation of common recurring manual operations in the creation of an extraction rule set from scratch. Using a historical rule set, we reconstruct

and describe the putative manual operations required to create it. In experiments targeting one key operation—the enumeration of words occurring in particular contexts—we simulate the process of corpus review and

word list creation, showing that several simple interventions greatly improve recall as a function of simulated labor.

**URL**

[*https://aclanthology.org/2022.pandl-1.6/*](https://aclanthology.org/2022.pandl-1.6/)

**File attachment**

*If the publication is not hosted externally (or you would like to include a download in addition to the URL), please provide the full quality PDF download.*

**Associated area of expertise**

*Speech & Natural Language / Information Extraction*