MODEL 1: (GEO + ML Methods)

pip install sense2vec

prodigy sense2vec.teach terms_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds "~~decision tree, decision trees, random forest, random forests, linear regression, regression, water-balance, water balance, aqueous model, subduction zone, species occurrence, transport model, sliding-block, nonlinear regression, chemical equilibrium, flow velocity, heat transport, linear sorption, Lagrangian transport, Hayes, bayesian, differential equation~~"

prodigy sense2vec.teach terms_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds "regression"

prodigy terms.to-patterns terms_method --label METHOD

prodigy db-out terms_method > /Users/sashaqanderson/Dropbox/USGS/NER_Work/method_patterns.jsonl

Prodigy ner.manual method_data blank:en /Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text.jsonl --label METHOD

prodigy db-out method_data > /Users/sashaqanderson/Dropbox/USGS/method_data_model.jsonl

python -m spacy pretrain /Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text.jsonl en_vectors_web_lg /Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_model2/ ~~--use-vectors~~

prodigy ner.batch-train method_data en_vectors_web_lg --init-tok2vec /Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_model2/model999.bin

Prodigy ner.batch-train method_data en_vectors_web_lg --init-tok2vec /Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_model2/model999.bin --output tmp_model --eval-split 0.2 --label METHOD

40% Accuracy

Prodigy ner.make-gold method_data_correct ./tmp_model /Users/sashaqanderson/Dropbox/USGS/NER_Work/ner_text.jsonl --label METHOD

Prodigy ner.batch-train method_data_correct en_vectors_web_lg --init-tok2vec /Users/sashaqanderson/Dropbox/USGS/NER_Work/pretrain_model2/model999.bin --output method_model --eval-split 0.2 --label METHOD --n-iter 20

64% Accuracy