ALGO MODEL COMMANDS for PRODIGY

prodigy sense2vec.teach algo_method /Users/sashaqanderson/Dropbox/USGS/NER_Work/s2v_old --seeds "logistic_regression, random_forest"

prodigy terms.to-patterns algo_method --label ALGO

{"label":"ALGO","pattern":[{"lower":"logistic_regression|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"random_forest|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"random_forests|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"genetic_algorithms|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"dynamic_programming|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"binary_search|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"Bayesian_approach|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"genetic_algorithm|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"clustering|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"gradient_descent|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"Bayesian_methods|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"decision_tree|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"Markov_chain|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"minimax|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"neural_networks|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"decision_trees|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"Bayesian_inference|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"neural_network|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"backprop|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"neural_nets|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"Bayesian_statistics|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"Dijkstra|PERSON"}]}
{"label":"ALGO","pattern":[{"lower":"neural_net|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"backpropagation|NOUN"}]}
{"label":"ALGO","pattern":[{"lower":"artificial_neural_networks|NOUN"}]}

prodigy db-out algo_method > /Users/sashaqanderson/Dropbox/USGS/NER_Work/arxiv_model/algo_patterns.jsonl

##

Prodigy ner.manual algo_data2 blank:en /Users/sashaqanderson/Dropbox/USGS/NER_Work/arxiv_model/arxiv_train.jsonl --label ALGO

prodigy db-out algo_data2 > /Users/sashaqanderson/Dropbox/USGS/NER_Work/arxiv_model/algo_data_model2.jsonl

python -m spacy pretrain

/Users/sashaqanderson/Dropbox/USGS/NER_Work/arxiv_model/arxiv_train.jsonl

en_vectors_web_lg /Users/sashaqanderson/Dropbox/USGS/NER_Work/arxiv_model/pretrain_algo_model/ --use-vectors

Prodigy ner.batch-train algo_data2 en_vectors_web_lg --init-tok2vec

/Users/sashaqanderson/Dropbox/USGS/NER_Work/arxiv_model/pretrain_algo_model/model999

.bin --output algo_model2 --eval-split 0.2 --label ALGO

25% Accuracy(improperly using the buttons)
46% (using the same pre-train)

Prodigy ner.make-gold algo_data_correct2 ./algo_model2
/Users/sashaqanderson/Dropbox/USGS/NER_Work/arxiv_model/arxiv_train.jsonl --label ALGO

Prodigy ner.batch-train algo_data_correct2 en_vectors_web_lg --init-tok2vec

/Users/sashaqanderson/Dropbox/USGS/NER_Work/arxiv_model/pretrain_algo_model/model999

.bin --output algo_model2 --eval-split 0.2 --label ALGO --n-iter 20

44% Start Fresh – I tried to use the other pre-trained model, but it was a bust

WAS USING THE IGNORE BUTTON INCORRECTLY – ONLY USE IF YOU DON"T KNOW THE ANSWER, OTHERWISE USE REJECT BUTTON – redo this test