sampling

↪ input

data/proteome_human.csv
data/words_hp.csv

output →

data/current_sequences.csv
data/current_words.csv

embedding_QSO

↪ input

data/current_sequences.csv

output →

postprocessing/QSO.csv

embedding_termfreq

↪ input

data/current_sequences.csv
data/current_words.csv

output →

postprocessing/termfreq.csv

embedding_seq2vec

↪ input

data/current_sequences.csv
data/current_words.csv
data/ids_hp_w5.csv
data/weights_w5_d100.csv

output →

postprocessing/seq2vec.csv

seq2vec_weighting

↪ input

data/TF_IDF.csv
data/current_sequences.csv
data/current_words.csv
data/ids_hp_w5.csv
data/weights_w5_d100.csv

output →

postprocessing/seq2vec_SIF.csv
postprocessing/seq2vec_TFIDF.csv

embedding_biophys

↪ input

data/current_sequences.csv

output →

postprocessing/biophys.csv

biophys_weighting

↪ input

data/TF_IDF.csv
data/current_sequences.csv
data/current_words.csv
data/ids_hp_w5.csv
data/weights_w5_d100.csv

output →

postprocessing/biophys_SIF.csv
postprocessing/biophys_TFIDF.csv

embedding_random

↪ input

data/current_sequences.csv

output →

postprocessing/random.csv

random_weighting

↪ input

data/TF_IDF.csv
data/current_sequences.csv
data/current_words.csv
data/ids_hp_w5.csv
data/weights_w5_d100.csv

output →

postprocessing/random_SIF.csv
postprocessing/random_TFIDF.csv

true_similarity_syntax

↪ input

data/current_sequences.csv

output →

similarity/true_syntax.csv

CCR

↪ input

postprocessing/QSO.csv
postprocessing/biophys.csv
postprocessing/biophys_SIF.csv
postprocessing/biophys_TFIDF.csv
postprocessing/random.csv
postprocessing/random_SIF.csv
postprocessing/random_TFIDF.csv
postprocessing/seq2vec.csv
postprocessing/seq2vec_SIF.csv
postprocessing/seq2vec_TFIDF.csv
postprocessing/termfreq.csv

output →

postprocessing/QSO_CCR.csv
postprocessing/biophys_CCR.csv
postprocessing/biophys_SIF_CCR.csv
postprocessing/biophys_TFIDF_CCR.csv
postprocessing/random_CCR.csv
postprocessing/random_SIF_CCR.csv
postprocessing/random_TFIDF_CCR.csv
postprocessing/seq2vec_CCR.csv
postprocessing/seq2vec_SIF_CCR.csv
postprocessing/seq2vec_TFIDF_CCR.csv
postprocessing/termfreq_CCR.csv

similarity

↪ input

postprocessing/QSO.csv
postprocessing/QSO_CCR.csv
postprocessing/biophys.csv
postprocessing/biophys_CCR.csv
postprocessing/biophys_SIF.csv
postprocessing/biophys_SIF_CCR.csv
postprocessing/biophys_TFIDF.csv
postprocessing/biophys_TFIDF_CCR.csv
postprocessing/random.csv
postprocessing/random_CCR.csv
postprocessing/random_SIF.csv
postprocessing/random_SIF_CCR.csv
postprocessing/random_TFIDF.csv
postprocessing/random_TFIDF_CCR.csv
postprocessing/seq2vec.csv
postprocessing/seq2vec_CCR.csv
postprocessing/seq2vec_SIF.csv
postprocessing/seq2vec_SIF_CCR.csv
postprocessing/seq2vec_TFIDF.csv
postprocessing/seq2vec_TFIDF_CCR.csv
postprocessing/termfreq.csv
postprocessing/termfreq_CCR.csv

output →

similarity/QSO.csv
similarity/QSO_CCR.csv
similarity/biophys.csv
similarity/biophys_CCR.csv
similarity/biophys_SIF.csv
similarity/biophys_SIF_CCR.csv
similarity/biophys_TFIDF.csv
similarity/biophys_TFIDF_CCR.csv
similarity/random.csv
similarity/random_CCR.csv
similarity/random_SIF.csv
similarity/random_SIF_CCR.csv
similarity/random_TFIDF.csv
similarity/random_TFIDF_CCR.csv
similarity/seq2vec.csv
similarity/seq2vec_CCR.csv
similarity/seq2vec_SIF.csv
similarity/seq2vec_SIF_CCR.csv
similarity/seq2vec_TFIDF.csv
similarity/seq2vec_TFIDF_CCR.csv
similarity/termfreq.csv
similarity/termfreq_CCR.csv

true_similarity_semantics

↪ input

data/current_sequences.csv

output →

similarity/true_semantics_BP.csv
similarity/true_semantics_CC.csv
similarity/true_semantics_MF.csv

evaluation

↪ input

similarity/QSO.csv
similarity/QSO_CCR.csv
similarity/biophys.csv
similarity/biophys_CCR.csv
similarity/biophys_SIF.csv
similarity/biophys_SIF_CCR.csv
similarity/biophys_TFIDF.csv
similarity/biophys_TFIDF_CCR.csv
similarity/random.csv
similarity/random_CCR.csv
similarity/random_SIF.csv
similarity/random_SIF_CCR.csv
similarity/random_TFIDF.csv
similarity/random_TFIDF_CCR.csv
similarity/seq2vec.csv
similarity/seq2vec_CCR.csv
similarity/seq2vec_SIF.csv
similarity/seq2vec_SIF_CCR.csv
similarity/seq2vec_TFIDF.csv
similarity/seq2vec_TFIDF_CCR.csv
similarity/termfreq.csv
similarity/termfreq_CCR.csv
similarity/true_semantics_BP.csv
similarity/true_semantics_CC.csv
similarity/true_semantics_MF.csv
similarity/true_syntax.csv

output →

similarity/scores/QSO.txt
similarity/scores/QSO_CCR.txt
similarity/scores/biophys.txt
similarity/scores/biophys_CCR.txt
similarity/scores/biophys_SIF.txt
similarity/scores/biophys_SIF_CCR.txt
similarity/scores/biophys_TFIDF.txt
similarity/scores/biophys_TFIDF_CCR.txt
similarity/scores/random.txt
similarity/scores/random_CCR.txt
similarity/scores/random_SIF.txt
similarity/scores/random_SIF_CCR.txt
similarity/scores/random_TFIDF.txt
similarity/scores/random_TFIDF_CCR.txt
similarity/scores/seq2vec.txt
similarity/scores/seq2vec_CCR.txt
similarity/scores/seq2vec_SIF.txt
similarity/scores/seq2vec_SIF_CCR.txt
similarity/scores/seq2vec_TFIDF.txt
similarity/scores/seq2vec_TFIDF_CCR.txt
similarity/scores/termfreq.txt
similarity/scores/termfreq_CCR.txt

downstream

↪ input

similarity/scores/QSO.txt
similarity/scores/QSO_CCR.txt
similarity/scores/biophys.txt
similarity/scores/biophys_CCR.txt
similarity/scores/biophys_SIF.txt
similarity/scores/biophys_SIF_CCR.txt
similarity/scores/biophys_TFIDF.txt
similarity/scores/biophys_TFIDF_CCR.txt
similarity/scores/random.txt
similarity/scores/random_CCR.txt
similarity/scores/random_SIF.txt
similarity/scores/random_SIF_CCR.txt
similarity/scores/random_TFIDF.txt
similarity/scores/random_TFIDF_CCR.txt
similarity/scores/seq2vec.txt
similarity/scores/seq2vec_CCR.txt
similarity/scores/seq2vec_SIF.txt
similarity/scores/seq2vec_SIF_CCR.txt
similarity/scores/seq2vec_TFIDF.txt
similarity/scores/seq2vec_TFIDF_CCR.txt
similarity/scores/termfreq.txt
similarity/scores/termfreq_CCR.txt

output →

scores_w5_d100.csv

all

↪ input

scores_w5_d100.csv
similarity/scores/QSO.txt
similarity/scores/QSO_CCR.txt
similarity/scores/biophys.txt
similarity/scores/biophys_CCR.txt
similarity/scores/biophys_SIF.txt
similarity/scores/biophys_SIF_CCR.txt
similarity/scores/biophys_TFIDF.txt
similarity/scores/biophys_TFIDF_CCR.txt
similarity/scores/random.txt
similarity/scores/random_CCR.txt
similarity/scores/random_SIF.txt
similarity/scores/random_SIF_CCR.txt
similarity/scores/random_TFIDF.txt
similarity/scores/random_TFIDF_CCR.txt
similarity/scores/seq2vec.txt
similarity/scores/seq2vec_CCR.txt
similarity/scores/seq2vec_SIF.txt
similarity/scores/seq2vec_SIF_CCR.txt
similarity/scores/seq2vec_TFIDF.txt
similarity/scores/seq2vec_TFIDF_CCR.txt
similarity/scores/termfreq.txt
similarity/scores/termfreq_CCR.txt