Project Documentation

RI Annotation Package :

* convert\_xml : utilities to transform the original xml documents into the format used for the annotation task, and prepare batches
* agreementEvaluation : functions to compute inter-annotator agreement
* main\_annotations : applies functions from both previous modules
* extract\_annotations : extracts i2b2 original xml annotations to create the data tables found in DataTables

SVM\_Anchoring package :

* filter\_ri\_timexes : performs the first filtering of relative and imprecise time expressions (before the annotation step)
* svm\_anchoring : contains all the functions to train svm classifier on the original anchor problem
* map\_custom\_annotations : utilities to transform our annotations into a format that can be used to train svm classifiers
* main\_normalization : executes functions from the other scripts

Bert\_Anchoring package :

* bert\_for\_multilabel\_classification : redefines the BertModel class to accept multi-label classification (where more than one label can be true)
* convert\_for\_bert : converts the annotated data to bert type inputs. The inputs are saved in dataframes to save time during training.
* dataset : a utility class to organise the different types of data
* bert\_transformer : the file which holds the functions used to train and test bert
* bert\_tgpu\_training\_and\_testing : run this file to train a new bert model on a gpu server, and test it
* test\_bert\_model : use this file to test a bert model that was previously saved