Figure 1 illustrate our method. We start with fetching disease annotation from omim ,generif and ensembl variation. This annotation data is not standardlize, i.e. plain texts are used to describe a disease/phenotype instead of ontology terms. We then use MetaMap and NCBO Annotator to recognize terms from an ontology of interest(e.g.,HDO) in the annotation. Next, we associate the recognized ontology terms with the gene identifiers and store them in our database. Finally, we calculate term enrichment for list of genes with this standardlized gene-disease annotation data, along with a domain specified clip tree.

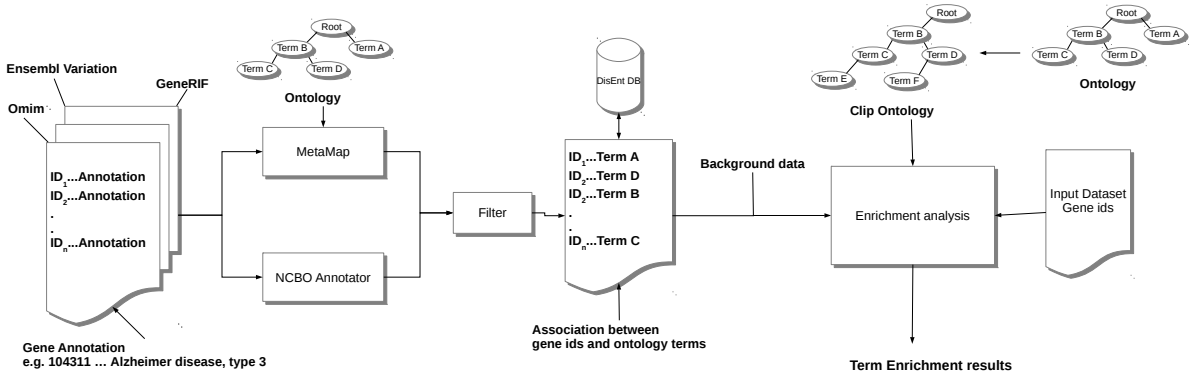

Figure 1: Workflow for generating the backgrand annotation sets for DisEnt tool. Gene-disease associations are acquired from omim,generif and ensembl variation. The annotations are usually raw text with no overall relations between each other. MetaMap and NCBO Annotator are then used to recognize terms from the annotation.