Step 1: Run ELVira - **sh makeElWithoutReasoning.sh -i hp-full.owl -o hp_elvira_8sep.owl**
-I -> input file
-o -> output file
It outputs HP Ontology(**hp-full.owl**) to reduced EL subset form(**hp_elvira_8sep.owl**).
Note1: How to install ELVira- https://code.google.com/archive/p/el-vira/downloads

Step 2: Some irrelevant axioms for our task may cause error while running normalizer code. So toi remove them run the code at **Preprocessing.ipynb** under **Removing Irrelevant Axioms**.
So from here we obtain **hp_elvira_8sep_trimmed.owl**.

Step 3: The reflexive axioms causes error in normalizer code. So change them as follows, in our owl file there is only one such case.

changes to

Reflexive_Axiom_Modifier.ipynb can be used for this purpose.

Step 4: hp_elvira_8sep_trimmed.owl is reduced to normalized form using:
groovy -cp jar/jcel.jar Normalizer.groovy -o
/home/aniket/Downloads/elvira-0.2/hp_8sep_norm.owl -i
/home/aniket/Downloads/elvira-0.2/hp_elvira_8sep_trimmed.owl

Note2: how to install normalizer module-
[https://github.com/bio-ontology-research-group/el-embeddings/tree/37ea9ded9e1b0eef6a c38632863bc7e4ead29708](https://github.com/bio-ontology-research-group/el-embeddings/tree/37ea9ded9e1b0eef6ac38632863bc7e4ead29708) **README.md**

Step 5: Run **AddRelations.ipynb** to integrate *hasAnnotation* and *Dir* axioms to ontology.
Change the input and output files as-
df_train – the train g-d pair file
norm_ontology- normalized ontology
annot_file – the annotation file
output_finalized_ontology- finalized ontology integrating all data that will be passed to ElEm model.

Step6: Run **Train_Test_Split.ipynb** to generate the 4 train-test sets
Change the input as:
pair file- the name of file that contains the g-d associations
n- number of instances each of positive and negative association you wish to see in test sets. If n=2046 then there will be 2046 positive and negative pairs in the test set I.e, 4912 total pairs.

Step 7: Run **ElEm_Algorithm.ipynb** to generate the embeddings.
Change input as
FILE_NAME – name of finalized ontology(.owl) file.
Follow the instructions given in the notebook.

Step 8: Run **Rank_Distribution_Plot.ipynb** to calculate rank distribution of each model.
Change input, output as:
test_data- g-d test file
all_pairs- full g-d pair
lists- lists of all the embedding files. Note the token for each file ('**ElEm','TransE1','rdf2vec'**) should be same.
rel_file- name of relation embedding model. Required only for ElEm model.
output_rank_distribution- the file where rank distribution will be saved.

Step9: Run **Evaluation.ipynb** to generate performance statistics of KGE models.
input_rank_file- rank distribution file from prev. step.
output_stats_file_name- output file where the statistics are saved.

Note: for other models like TransE, OWL2Vec etc.,- refer-
https://github.com/liseda-lab/KGE_Predictions_GD