**Data preprocessing for TKRL**

---

All original data could be found in data.rar or be downloaded through links in README.md, while users should generate the following files according to their own tasks:

../data/entity2id.txt
../data/relation2id.txt
../data/type2id.txt
../data/domain2id.txt
../data/relationType.txt
../data/relationDomain.txt
../data/typeEntity.txt

The datasets are extracted from Freebase, and the hierarchical type structure we use is as follows:
/domain/type/topic
e.g. /book/author/William Shakespeare
/book is the domain, while /book/author is the type. William Shakespeare stands for the entity.

---

**../data/entity2id.txt**
entity list, the format is:
[Name id]

**../data/relation2id.txt**
relation list, the format is:
[Name id]

**../data/type2id.txt**
type list, the format is:
[Name id]

**../data/domain2id.txt**
domain list, the format is:
[Name id]

**../data/relationType.txt**
relation-specific information, indicating the correct type head/tail should belong to in a specific relation, the format is:
[relation type of head type of tail]

**../data/relationDomain.txt**
relation-specific information, indicating the correct domain head/tail should belong to in a specific

relation, the format is:
[relation domain of head domain of tail]

**../data/typeEntity.txt**

Indicating all entities which could belong to a type, for soft type constraints in training and evaluation, the format is:
[type_id entity_id_1 … entity_id_n]