Freebase

(Freebase-triples: A Methodology for Processing the Freebase Data Dumps[1])

1. Freebase Schema

- Topic: a distinct entity or object
- Mid: a unique machineId identifier, in the format "/m/+{alphanumeric}" (e.g. /m/abc123)
- Type: Person type is written in a format as /people/person
- Property: each type's further granular data. A person's Place of Birth is written in a format as /people/person/place_of_birth
- Triples format: (subject, predicate, object), such as
  (/m/abc123, /type/object/type, /people/person)
  (/m/abc123, /people/person/place_of_birth, /m/clnclna771)
- Forward slashes: /domain/type/property (e.g. /people/person/place_of_birth)

2. Freebase Data Dumps (1.9 billion triples)

- Changes:
  1) Slashes->full stop: /m/abc123->/m.abc123; /people/person ->/people.person
  2) All objects are encoded with a full URL path for global compatibility, such as http://rdf.freebase.com/ns/m.abc123
- Structure: each line is shown as follows
  <http://rdf.freebase.com/ns/g.112ygbz6>"\t"
  < http://rdf.freebase.com/ns/type.object/type.object.type>"\t"
  < http://rdf.freebase.com/ns/film.film>".\n"

FB15K

(Translating Embeddings for Modeling Multi-relational Data[2])

- 592,213 triplets with 14,951 entities and 1,345 relationships
- Select entities in Freebase that are also present in the Wikilinks database and that also have at least 100 mentions in Freebase (for both entities and relationships).

FB15K+

(Representation Learning of Knowledge Graphs with Hierarchical Types[3])

The same entities in FB15K but almost all relations between those entities, only discard relations which just appear once because of the sparsity in real-world KGs.

https://github.com/thunlp/TKRL 该项目对 FB15K 数据集的整理

File	Format
entity2id.txt	[Name id]
relation2id.txt	[Name id]
type2id.txt	[Name id]
domain2id.txt	[Name id]
relaitonType.txt	[relation type_of_head type_of_tail]
relationDomain.txt	[relation domain_of_head domain_of_tail]
typeEntity.txt	[type_id entity_id_1 ··· entity_id_n]
entity2type	[entity type_1 ··· type_n]
relation_specific	[relation head_type tail_type]

MMKG

(MMKG: Multi-Modal Knowledge Graphs[4])

FB15K as start point, create DB15K and YAGO15K.

1. Extract alignments between entities of FB15K and DBPedia via sameAs relation
2. Include additional entities to make the two kgs have the same number of entities and to have unaligned entities
3. Numerical literals
4. Image:
    1) Extract all Wikipedia URIs from Freebase Data Dumps for disambiguation
    2) Entity name, entity name followed by type, Wikipedia URIs as query strings
    3) Web crawler on Google Images, Bing Images and Yahoo Image Search
    4) Store 20 top ranked images retrieved by each browser
    5) Filter by size and quality (corrupted, low quality and duplicate)
    6) Scale the images while maintaining their aspect ratio
    7) Distribute a distinct image to FB15K and DB15K

参考文献

[1] Chah N . Freebase-triples: A Methodology for Processing the Freebase Data Dumps[J]. 2017.

[2] Bordes A , Usunier N , Garcia-Duran A , et al. Translating Embeddings for Modeling Multi-relational Data. Curran Associates Inc. 2013.

[3] Xie R , Liu Z , Sun M . Representation learning of knowledge graphs with hierarchical types. AAAI Press, 2016.

[4] Liu Y , Li H , Garcia-Duran A , et al. MMKG: Multi-Modal Knowledge Graphs[C]// Springer, Cham. Springer, Cham, 2019.