1. 华东

<http://openkg.cn/dataset/symptom-in-chinese>

| **域** | **价值** |  |
| --- | --- | --- |
| **源** | [http://dsc.nlp-bigdatalab.org:8081/search.html](http://dsc.nlp-bigdatalab.org:8081/search.html" \t "http://openkg.cn/dataset/_blank) |  |
| **作者** | [华东理工大学](mailto:wmj_ecust@163.com) |  |
| **维护者** | [王梦婕](mailto:wmj_ecust@163.com) |  |
| **最近更新** | 2018年3月26日, 下午3点14分 (UTC+08:00) |  |
| **创建的** | 2016年7月12日, 下午3点37分 (UTC+08:00) |  |
| **三元组数量** | 617,499 | 备注：617,462 |
| **实体数量** | 135,485 | 备注：73,797 |
| **备注：关系数量** |  | 备注：18 |

注1：dsc.nlp-bigdatalab.org.ttl文件【个人统计数字（备注）与官网给的数量不一致】

line1-line5定义了缩写，不计入三元组

line7-line75定义了schema（含ontology），不计入三元组

其余行，三元组第二个元素含'rdf-schema#label'，是实体名字，不计入三元组；含'rdf-syntax-ns#type'，是语义网的is-a关系，不计入三元组；含'owl:sameAs'（line969955之后），是本数据和umls数据对应关系，不计入三元组；含'rdfs:label': （line976004以后），是umls数据的实体名，不计入三元组

注2：“数据集质量评估结果”中含有三份抽样评价

其中涉及umls的问题：由于umls更新过id，因此对应的umls数据名字是对的，id是错的，需要更新。

1. 39健康网疾病百科

<http://jbk.39.net/bw/t1/>，以疾病、症状、检查、手术组织，结构化程度较高，附件有对疾病类、症状类的爬虫（症状类和手术类解析未完成）。

1. Umls

<http://www.linkedlifedata.com/about>

集合了25个数据源，但是英文的，直接使用难度较大。

爬取较难，id取值范围未知，如下网址是以fever为object的三元组示例

<http://linkedlifedata.com/resource/umls-concept/C0015967?generic=true&role=object&context=http%3A%2F%2Flinkedlifedata.com%2Fresource%2Fmappings&_ajax=selection+results>

1. 脑网络 <http://www.linked-brain-data.org/>

该图谱仅提供了相关关系（脑结构-认知功能；脑结构-脑疾病；认知功能-脑疾病），但是有条件概率统计。

该图谱重在精准医疗方面，对脑结构细分为基因、蛋白质等五个层次，每个层次之间也有关联关系。

其他的如中医药学知识图谱（知识歧义太多，歧义太多，http://openkg.cn/dataset/tcmsn）