## Bacterial biotope赛道

<https://sites.google.com/view/bb-2019/home>

识别科学文本和教科书中出现的微生物及其生境、表型的实体概念，利用领域知识资源（分类学和**本体论**）对实体进行标准化，并对实体进行关系提取。

四种实体类型：

-**microorgannisms** 微生物

-**habitats** 栖息地

-**geographical places** 地理环境

-**phenotypes** 表现型

两种关系类型：

**live\_in** 连接 microorgannisms 和 地点（habitats 或 geographical places）

**exhibits**连接 microorgan­nisms和 phenotypes

microorgannisms利用NCBI分类词典做标准化，

habitats和phenotypes利用OntoBiotope做标准化



Corpus:

PubMed数据库中微生物相关论文的标题及摘要

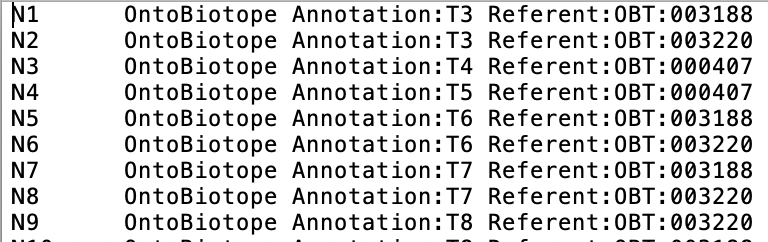
食品领域微生物相关文章内容摘取

标注是BioNLP-ST标注格式

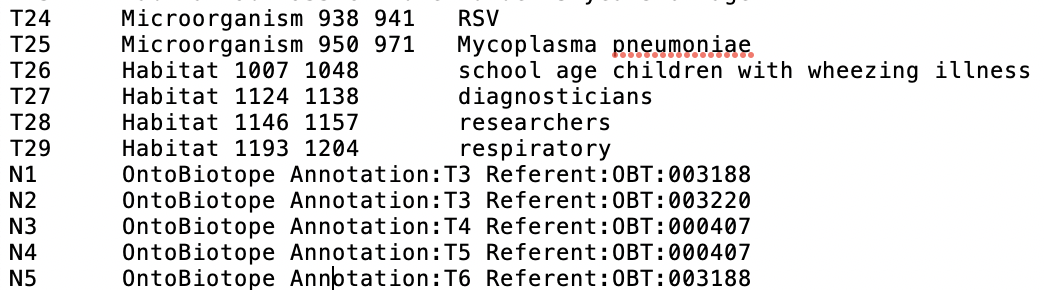
<https://drive.google.com/file/d/1G0po_xlRjQCZ-qxuA_4PLdipXU6rtYTp/view?usp=sharing>

### 任务1.

BB-norm: 对microorgannisms，habitats 和 phenotypes进行标准化。提供实体标注。



BB-norm+ner: 识别microorgannisms，habitats 和 phenotypes实体并进行标准化。

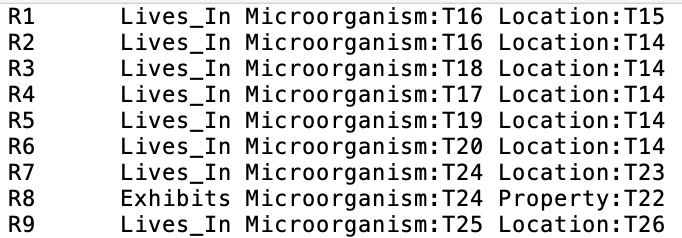


评价标准：类别预测的准确度。概念距离。如果实体对应多个类别，需要预测所有可能的类别。

For norm+ner, boundary accuracy will be factored in the evaluation since the inclusion or exclusion of modifiers can change the meaning and the normalization of phrases.

### 任务2.

BB-rel: 识别microorgannisms和habitats 或 geographical places之间的live\_in关系，microorgannisms和 phenotypes的exhibits关系。 提供实体标注（是否标准化？否）。



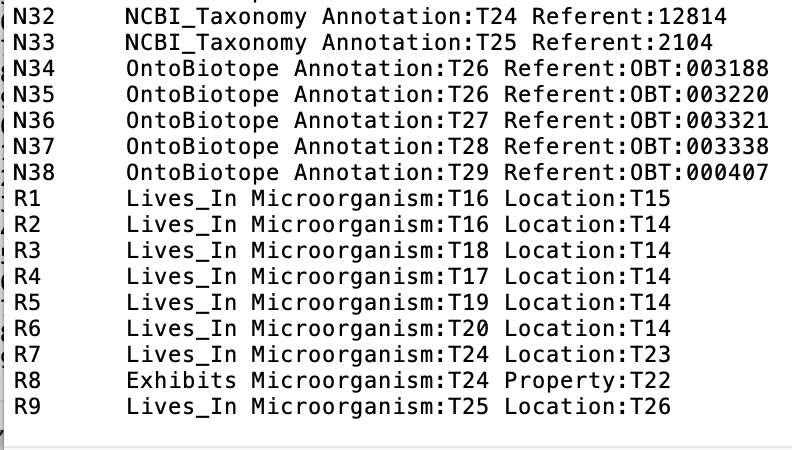
BB-rel+ner:识别实体并进行关系抽取。

评价标准：Recall+Precision

For rel+ner, boundary accuracy will be factored in the evaluation.

### 任务3.

建立知识库：前两个任务的集合。目标是衡量能够从文本中提取出多少信息。结果为文本层面。



评价标准：Recall+Precision

For rel+ner, boundary accuracy will be factored in the evaluation.