### 语义增：

例如给出这样一个问题，show the information about obama。针对这个问题，gAnswer以及目前存在的所有的将N转化为Q的框架都无法解决。

问题在于”Obama”这个实体可以在数据集中寻找到对应值。但是information不行。Information是一个抽象概念，也就是说这个information类似于一个类，其中包含许多概念，例如在此处information可能包括（age, wife, father, country….），这一类问题目前是无法解决的。而这种类型的问题又很常见，例如，奥巴马的联系方式（电话,email ，wechat, 推特…）

**解决思路：**

**首先定位问题。**

1. 当一个N中只含有一个实体的情况下执行语义增。
2. 当N中含有多个实体，但是其中存在一部分结构只存在一个实体，而且这部分结构会修饰另外一个实体的时候，执行语义增。

|  |  |
| --- | --- |
| **根据统计信息** | **先补后排序** |
| 统计信息很好理解，通过提前进行数据分析，构造一个类似的<key,value>表。例如key为联系方式，value为tel ,email … | 首先，查找出所有与这个实体相关的谓词集合Ps，之后在这个集合中根据N中的其他辅助信息进行谓词的选择和重要度排序。（未解决） |

### 语义改：

例如存在这样的一个问题，who is Obama’s father’s father？其生成的sparql查询为

Select ?name

{ ?x name Obama. ?x father ?y . ?y father ?name. }

但是事实上，这一步完全是可以简化的。因为father’s father与grandfather是等价的。因此这个查询Q完全可以等价为以下这个查询。

Select ?name

{ ?x name Obama. ?x grandfather ?name. }

这一步，采用本体文件中的信息来进行等价代换。

### 语义删：

这一步最好理解，主要目的为只保留句法分析树中的核心实体，将无意义词语，以及对生成Q 贡献不大的词语进行删除。

**有数据集如下：**

|  |  |
| --- | --- |
| **N** | **准确Q** |
| **who is Obama’s father?** | **Q1** |
| **Obama’s father?** | **Q1** |
| **Obama father?** | **Q1** |
| **who is it or and …** | **null** |

将N中所有的单词向量化为Wn。之后将N向量化为Vn。Vn为N中所有的Wn之和。

使用trans系列将Q向量化为Qv。

之后进行训练，保持Vn = Qv.

整个数据集训练完毕后，得到了所有的Wn。之后将Wn中所有维度值相加，其和为这个单词的核心度。因此可以根据核心度大小来选择删除一些单词，来提高效率。