

读书报告--语义 parser

1. 个人对 parser 的见解

假设现在我们进入一个未知的领域，需要掌握一门未知的语言，而现在我们知道这门未知的语言有着和我们熟知的语言（英语或者汉语）一样的 POS（即动词，名词，形容词，副词，介词等等），如下所示：

Amj_DT sda_NN dss_IN Alpha_NN gagerds_VB fdafe_NN fdsg_IN fdfds_DT dfk_JJ fdga_NN jik_IN sdf_NN.

(上面 NN 是名词，VB 是动词，IN 是介词，DT 是冠词，JJ 是形容词)

当我们看到这个的时候，第一反应肯定是不知所措，定下心来细细思考之后可能会有一些来自灵感，DT 和 NN 是可以先组合的，DT_JJ_NN 也是可以先组合的，于是我们就可以得到下面的例子：

Amj sda_(DT+NN) dss_IN Alpha_NN gagerds_VB fdafe_NN fdsg_IN fdfds dfk fdga_(DT+JJ+NN) jik_IN sdf_NN.

然后一个精通语言的 oracle 出来告诉我们，其实 Amj+sda+dss 是一个固定词组，它相当于 xxx_JJ，于是我们就可以把上面的句子进一步转化为：

Amj sda_(DT+NN) dss_IN Alpha_NN gagerds_VB fdafe_NN fdsg_IN fdfds dfk fdga_(DT+JJ+NN) jik_IN sdf_NN.

-->

Amj sda dss Alpha_(JJ+NN) gagerds_VB fdafe_NN fdsg_IN fdfds dfk fdga_(DT+JJ+NN) jik_IN sdf_NN.

形式上看上去简洁了很多，但是还有一个非常重要的问题没有解决，就是传说中的搭配问题，而这里面最重要的一个就是介词短语是修饰哪一个词的问题。

现在精通语言的 oracle 又给了我们一个提示，在这里，有一张表

fdfds dfk fdga 可以用来修饰 fdsg

fdfds dfk fdga 可以用来修饰 gagerds

sdf 可以用来修饰 fdfds dfk fdga_

这个就像我们常见的 I see a man with a telescope. 一样，我们可以理解为：

① I see a man. The man is with a telescope.

② I see a man by using a telescope.

也就是说电脑有的只是一些字符串，但是它却不知道是什么意思，如果要做出一个 dependency parser，我们就需要给与这些单词之间的联系，传统的基于统计的方法可以统计出一些信息，但是毫无疑问这是需要训练的，而且训练的话都是基于经过专家标注的修饰关系的语料库进行的，没有在语料库出现的基本是无法识别的。与其一直在同一语料库上尝试不同的算法，还不如直接人工把这些信息传给计算机。