标点引擎数据结构设计

条件随机场使用crf++开源库实现，下面对主要利用的函数进行说明。

主要接口和库函数为：Punctuation.h、libcrflearn.so、libPunctuation.so

# CRFPP::Model类

主要方法：

createModel(const char \*arg)

主要功能：初始化模型

参数：模型文件名及相关参数，例如“-v1 -m model”,-v1指定详细信息的层级，可以输出模型的详细信息，这里使用“-v1”即可。

createTagger()

主要功能：由model调用，生成Tagger实例，多线调用时共享模型

# CRFPP::Tagger类

主要方法：

bool add(const char\* str)

主要功能：将分词添加到当前预测序列文本中

参数：分词

bool parse()

函数功能：解析文本序列，为文本序列添加标记

size\_t xsize()

函数功能：返回列的大小

size\_t ysize()

函数功能：返回输出tags的大小

const char\* y2(size\_t i)

函数功能：返回第i-th token的输出标签

bool clear()

函数功能：清除所有上下文token