# Kuwo加tag的后处理逻辑

对于Bilstm-CRF模型**预测的tag结果**的termList和**原始query**两个维度进行依据**规则**的分析.

增加query的维度，主要由于模型分词错误导致无法分析出有效信息，如*“query=必点的热门的歌曲”*在termList中分词可能就错了，因此上一步结果无法参考。

大致步骤如下：

* 对于模型预测的结果tagList(termList)主要进行两步操作
  + term去不必要结尾
  + 包含标志词的tag=>"show"
* 对于query主要做"addscore"的处理
* 最终将以上两部分结果append到tokenRes，return

具体过程如下：

tag\_reg = re.compile(r'最热|很热|最新|最火|很火|经典|好听|热门|必点|歌曲|音乐|火热|超级')  
err\_reg = re.compile(r'最近|演奏|mp3|欣赏|播放|连播|连续|到哭|有没|又的|很快入|歌星|推荐|300|100|5000|应该|上很|一般'  
termList,tagList = Kuwo\_TagModel.predict(query)  
  
# STEP1. 对于Bilstm-CRF模型预测的tag结果的termList进行分析  
for index in range(0,len(termList)): # 对于每一个（term，tag）  
 if tagList[index] == "tag": # 作为query分析的加权flag  
 hastag = 1  
 # 处理term  
 对于tagList[index]="tag" 且为以下情况时，删除不必要结尾词：  
 if termList[index]以“又”、“上”、“时”结尾 or (termList[index].endswith("欣") and termList[index+1].startswith("赏")):  
 删除termList[index] # 删除不必要结尾词  
 if termList[index]以“的时候”结尾：  
 删除termList[index：index+3] # 删除不必要结尾词   
 if termList[index]以“时候”结尾：  
 删除termList[index：index+2] # 删除不必要结尾词  
 # 处理tag  
 对于tagList[index]="tag" 且为以下情况时，标记为show类型tag：  
 if termList[index]包含err\_reg词 or len(termList[index])<2:  
 tagList[index] = "show"  
 对于tagList[index]=="tag" and termList[index]包含tag\_reg词：   
 tagList[index] = "show"  
 将termList[index]、tagList[index]封装进tmpDic(term,final\_tag&tags)  
 tokenRes.append(tmpDic) # add term的分析结果  
  
# STEP2. 对于原始query进行分析  
# 利用query直接匹配 对于“query=必点的热门的歌曲”在termList中分词可能就错了，因此上一步结果无法参考  
res\_lis = tag\_reg.findall(query)   
if query包含tag\_reg词语：  
 for item in res\_lis:  
 if hastag == 0： #之前已有tag  
 tmpDic["final\_tag&tags"] = "tag"  
 else:  
 tmpDic["final\_tag&tags"] = "addscore" # 加权   
 tokenRes.append(tmpDic) # add term的分析结果  
else：  
 if hastag == 0： #之前已有tag  
 tmpDic["final\_tag&tags"] = "tag"   
 tmpDic["term"] = "歌"  
 tokenRes.append(tmpDic) # add term的分析结果  
  
# STEP3. 返回  
return tokenRes