courpus\_V3.json 為HSMM模型所產出的模板。

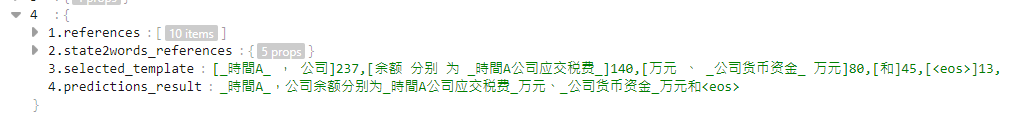
(其他.json 為早期訓練出的版本，格式也皆相同)

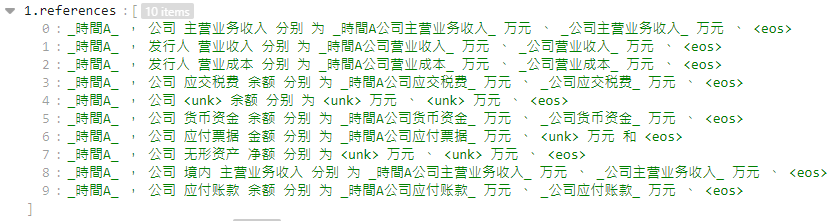
詳細規格如下:

以courpus\_V3.json為例

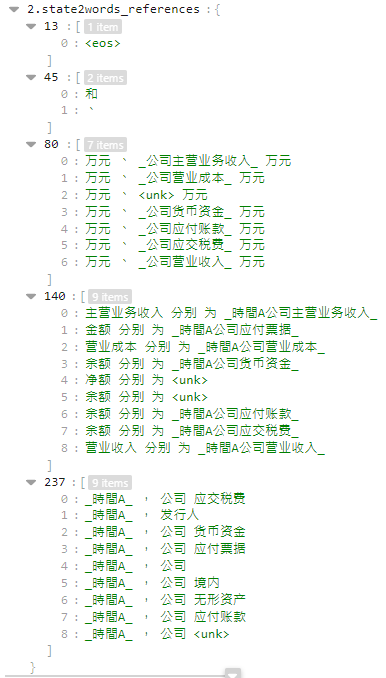
內共計4560個樣式模板



每個樣式的模板內有1.references、2.state2words\_references、3.selected\_template、4.predictions\_result 四樣資訊

1.references 為該樣式模板中，模型有參考原數據集的哪些句子，通常以此資訊來判定模型是否有把相似句構、相似敘述內容的句子分群在一塊。

2.state2words\_references 為模型將內些句子排列分成各個詞狀態，每個詞狀態也都會將相似字詞、句構分群在同狀態下。以下圖為例，共有237、140、80、45、13 這幾個詞狀態。



3.selected\_template 表示此樣式模板為:

詞狀態237->詞狀態140->詞狀態80->詞狀態45->詞狀態13

此順序組成的

4.predictions\_result 為從每個詞狀態挑選字詞後，即可組合成一段句子。

根據所選的字詞不同，可組合成不同敘述的句子。